

Assessment of Twentieth-Century Regional Surface Temperature Trends using the GFDL CM2 Coupled Models

T. R. Knutson¹, T. L. Delworth¹, K. W. Dixon¹, I. M. Held¹, J. Lu²,
V. Ramaswamy¹, D. Schwarzkopf¹, G. Stenchikov³, and R. J. Stouffer¹

¹Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, New Jersey

²UCAR Visiting Scientist at GFDL/NOAA, Princeton, New Jersey

³Rutgers University, New Brunswick, New Jersey

Corresponding author:

Thomas R. Knutson

GFDL/NOAA

Forrestal Campus, U.S. Rt. 1N

Princeton, NJ 08542

Email: Tom.Knutson@noaa.gov

May 23, 2005

Submitted to *Journal of Climate* -- Apr. 30, 2005

Status: corrected region-definition map figure 8.

Abstract.

Historical climate simulations of the period 1861-2000 using two new GFDL global climate models (CM2.0 and CM2.1) are compared with observed surface temperatures. “All-forcing” runs include the effects of changes in well-mixed greenhouse gases, ozone, sulfates, black and organic carbon, volcanic aerosols, solar flux, and land-cover. Indirect effects of tropospheric aerosols on clouds and precipitation processes are not included. Ensembles of size 3 (CM2.0) and 5 (CM2.1) with all forcings are analyzed, along with smaller ensembles of natural-only and anthropogenic-only forcing, and multi-century control runs with no external forcing.

Observed warming trends on the global scale and in many regions are simulated fairly realistically in the all-forcing and anthropogenic-only forcing runs. In the all-forcing runs, the ensemble mean simulated warming over 1949-2000 is significantly smaller than observed over Northern Asia, Canada, and much of the Indian Ocean, but greater than observed over parts of the tropical North Atlantic, tropical central Pacific, and South America. Differences in Arctic Oscillation behavior between model and observation contribute substantially to the Northern Asia temperature discrepancies. In the all-forcing and natural-only runs, a temporary global cooling in the models during the 1880s not evident in the observed temperature records is volcanically forced. El Niño interactions complicate comparisons of observed and simulated temperature records for the El Chichón and Pinatubo eruptions during the early 1980s and early 1990s.

The simulations support previous findings that 20th century global warming has resulted from a combination of natural and anthropogenic forcing, with anthropogenic forcing being the dominant cause of the pronounced late-20th century warming. The regional results provide evidence for an emergent anthropogenic warming signal over many, if not most, regions of the globe. The warming signal has emerged rather monotonically in the Indian Ocean/Western Pacific warm pool during the past half century. The tropical and subtropical North Atlantic and the tropical eastern Pacific are examples of regions where the anthropogenic warming signal now appears to be emerging from a background of more substantial multi-decadal variability.

1. Introduction

The earth’s mean surface temperature has warmed by ~0.8°C over the past century according to historical temperature records, supported by other climate observations (Folland et al. 2001). The purpose of the present study is to compare observed surface temperature changes to those produced by two versions of a new global climate model forced by estimated historical changes in a number of climate forcing agents. Such comparisons assess our ability to interpret past climate variations in terms of known climate forcing agents based on our physical understanding of the earth’s climate system as embodied in the coupled climate models.

In this study, we use GFDL’s new CM2 coupled climate models (CM2.0 and CM2.1, Delworth et al. 2005) to simulate surface temperature variations over the period 1861-2000. The climate forcing agents included in the simulations include changes in well-mixed greenhouse gases, ozone, anthropogenic aerosols (direct effect only), solar irradiance, land-cover type, and volcanic aerosols. Ensembles of “all-forcing” experiments are run for both

CM2.0 (n=3) and CM2.1 (n=5) coupled models, along with additional natural-only and anthropogenic-only forcing experiments (n=1 for CM2.0, n=3 for CM2.1). Multi-century control integrations without external forcing are used to assess internal climate variability and reduce impacts of any model drifts unrelated to external forcing changes.

Global climate model historical experiments using similar sets of anthropogenic and natural forcings (~1860-2000) have been performed previously (e.g., Tett et al. 1999; Stott et al. 2000; Broccoli et al. 2003; Meehl et al. 2004; Hansen et al. 2005). These studies found that both natural and anthropogenic forcings made significant contributions to early 20th century surface temperature changes, with anthropogenic forcings (e.g., greenhouse gases) being the dominant cause of the warming in the second half of the 20th century. Delworth and Knutson (2000) found that internal climate variability also could have played a substantial role in the observed early 20th century warming.

The present study represents a substantial step beyond previous work at GFDL on this topic (e.g., Knutson et al., 1999; Delworth and Knutson 2000, Broccoli et al. 2003) in two important respects. First, we use two versions of a completely updated global coupled climate model developed over a period of several years at GFDL (Delworth et al. 2005). The new models are substantially improved over previous GFDL coupled climate models in that they provide much improved simulations of El Niño variability (Wittenberg et al. 2005), and they do not require the use of flux adjustments to control model drift. The new models include more realistic treatments of sea ice, land-surface processes, and other physical processes in the atmosphere and oceans (Delworth et al. 2005 and references therein) than the previous models. Second, the experiments in this paper include a more comprehensive and physically based set of climate forcing agents than our previous studies.

For comparing model-simulated and observed temperature variations, one has the choice of using multivariate “fingerprint” detection and attribution techniques (e.g., Hegerl et al. 1997) or simpler univariate methods based on local trends (e.g., Knutson et al. 1999; Boer et al. 2000) or climate indices (e.g., Karoly et al. 2003). A review of various methodologies and conclusions obtained through their application to previous coupled models is provided by Mitchell et al. (2001) in Chapter 12 of the Intergovernmental Panel on Climate Change (IPCC) Working Group 1 Third Assessment Report (TAR); see also Knutson et al. 1999). Multivariate methods enhance one’s chances of detecting and attributing climate change, while a univariate analysis of the significance of local trends retains the advantage of being relatively easy to understand and communicate to non-specialists. In addition to the presentation of trend maps and their significance, as in Knutson, et al. (1999) we have also included an extensive set of figures showing time series “case studies” for a number of regions of interest. This approach avoids the dependence of the trends on the selection of starting/ending dates. Concerning regional-scale analysis, recent studies have applied formal detection/attribution methods at decreasing spatial scales and report evidence for detectable anthropogenic warming signals down to the continental scale (Zwiers and Zhang, 2003; Stott 2003). Finally, while we focus on surface temperature changes using a single pair of models in this report, a number of studies (not reviewed here) have begun to assess 20th century climate changes using multi-model ensembles and climate variables other than surface temperature.

The paper is structured as follows. In sections 2 and 3, we present a brief overview of the main characteristics of the models and the climate forcing agents used in the historical simulations. In section 4, characteristics of the model control runs (without changes in

climate forcings) are examined. In section 5, we examine changes in global mean temperature in the historical forcing runs. In section 6, we compare the simulated and observed surface temperature changes on a regional basis. Arctic Oscillation influences are investigated in Sections 7. Section 8 contains our summary and conclusions.

2. Model Description

The two coupled models used for the present study (CM2.0 and CM2.1) are described in detail in Delworth et al. 2005) and references therein. Further information is available online at: <http://nomads.gfdl.noaa.gov/CM2.X/references/>, and model output data is freely available at: <http://nomads.gfdl.noaa.gov/>. These coupled models are composed of four component models: atmosphere, land, sea ice, and ocean. The coupling between the component models (Balaji et al. 2005) occurs at one and two hour intervals in CM2.0 and CM2.1, respectively.

The atmospheric model has a grid spacing of 2.5° longitude by 2° latitude and 24 vertical levels. The model contains a completely updated suite of model physics compared to the previous GFDL climate model, including new cloud prediction and boundary layer schemes, and diurnally varying solar insolation. The radiation code allows for explicit treatment of numerous radiatively important trace gases (including tropospheric and stratospheric ozone, halocarbons, etc.), a variety of natural and anthropogenic aerosols (including black carbon, organic carbon, tropospheric sulfate aerosols, and volcanic aerosols), and dust particles. Aerosols in the model do not interact with the cloud scheme, so that indirect aerosol effects on climate are not considered. A full description of the atmospheric model is contained in GFDL_GAMDT (2004) with updates as described in Delworth et al. (2005). CM2.1 and CM2.0 differ in several aspects including different dynamical cores (finite volume (Lin, 2004) in CM2.1 vs. a “B-grid” finite difference dynamical core in CM2.0), further tuning of the cloud scheme for CM2.1, and a modified formulation of evaporation from land grid points with frozen soil (see Delworth et al. 2005).

The land model used in both CM2.0 and CM2.1 is the Land Dynamics model (LaD) as described in Milly and Shmakin (2002). Surface water is routed instantaneously to ocean destination points on the basis of specified drainage basins. The land cover type in the model uses a classification scheme with 10 different land cover types.

The ocean model (Gnanadesikan et al. 2005; Griffies et al. 2005) has a nominal grid spacing of 1° in latitude and longitude, with meridional grid spacing decreasing in the tropics to $1/3^\circ$ near the equator, and uses a tripolar grid to avoid polar filtering over the Arctic. The model has 50 vertical levels, including 22 levels with 10 m thickness each in the top 220 m. A novel aspect is the use of a true fresh-water-flux boundary condition. The ocean components of CM2.0 and CM2.1 differ in parameter settings for some subgrid scale physics and time-stepping (Delworth et al. 2005).

The sea ice model, identical in the two models, is a dynamical model with three vertical layers and five ice thickness categories. The model uses the elastic-viscous-plastic rheology to calculate ice internal stresses, and a modified Semtner three-layer scheme for thermodynamics (Winton 2000).

For comparison to observed surface temperatures, the HadCRUT2v dataset (<http://www.cru.uea.ac.uk/cru/data/temperature/>) is used. This data set (1870-2003) combines the land surface air temperature data of Jones and Moberg (2003) with the HadSST1 sea surface temperature (SST) data of Parker et al. (1995) and Rayner et al. (2003),

where variance adjustments have been applied to both land and ocean data (Jones et al. 2001). Trends in this study refer to simple linear trends computed using least-squares regression.

3. Climate Forcings for Historical Runs

For the historical forcing simulations described in this report, the models were integrated for a period of 140 years using a time-varying specification of various climate forcing agents representative of conditions from 1861-2000. An ensemble of such integrations was created for each model, using initial conditions obtained from widely separated (40 years +) points in multi-century control integrations with fixed 1860 radiative forcing and land surface conditions.

The time-varying forcing agents included CO₂, CH₄, N₂O, halons, tropospheric and stratospheric O₃, anthropogenic tropospheric sulfates, black and organic carbon, volcanic aerosols, solar irradiance, and the distribution of land cover types. The time variations were based on a combination of observations and reconstructions for the late 19th and 20th centuries. A brief description of the land cover, solar, and volcanic aerosol data sets is given below. A comprehensive discussion of the data sources for the time variations and other details of all the forcings and their uncertainties are provided in Ramaswamy et al. (in preparation; see also http://data1.gfdl.noaa.gov/CM2.X/faq/question_13.html and <http://nomads.gfdl.noaa.gov/CM2.X/references/>).

The land cover-type change history is based on the Hurtt et al. (2005) global land use reconstruction history. Changes in land cover type affect the model's surface albedo, surface roughness, stomatal resistance, and effective water capacity. Solar variations over the period 1882-2000 are implemented as a function of wavelength, using data provided by J. Lean (Lean et al., 1995; Lean, personal communication, 2003; see also IPCC, 2001).

The volcanic aerosol data set was developed based on volcanic aerosol optical depth for visible band compiled from different sources by Sato et al. (1993). This data set was subsequently improved by Hansen et al. (2002) and column average zonal mean effective radius was provided for the entire period. During the satellite era the aerosol characteristics are mostly based on SAGE (Stratospheric Aerosol and Gas Experiment) observations.

4. Control Simulations

Multi-century control runs of both the CM2.0 and CM2.1 models, with constant radiative forcing agents and land cover type appropriate to 1860 conditions, were performed as a preliminary step. Details of the initialization of these 1860 control runs are given in Delworth et al. (2005) and are similar to that described in Stouffer et al. (2004). The control runs provide an estimate of the climate model's internal variability (variability in the absence of external forcing changes) and an estimate of any long-term drift in the model that results from the coupled model's climate not being fully in equilibrium with the 1860 forcing. The long control runs also provide widely separated initial conditions (particularly three-dimensional ocean initial conditions) to use to initiate independent ensemble members for the historical forcing scenarios. This statistical sampling of control run ocean initial conditions is necessary because the true three-dimensional state of the ocean from 1860 is poorly constrained by observations. A separate pair of control runs with constant 1990 climate forcings was also completed, and are used in this paper only for the maps comparing

observed and simulated local interannual variability. The 1990 control runs were used for this comparison since the observations are from this general time period.

a. Global mean temperature variability

Figure 1 summarizes a number of aspects of the CM2.0 and CM2.1 control runs used in the study in terms of global mean reference atmospheric temperature at 2 m above the surface (T_{ref}). The CM2.0 control run (Fig. 1 a) consists of 500 years of simulation, which was preceded by a 300-year “spin-up” period (Delworth et al. 2005) of which the last 90 years is shown in Fig. 1 (a). The time series of T_{ref} shows that the CM2.0 model continues to warm at a rate of about $0.2^{\circ}\text{C } 100 \text{ yr}^{-1}$ for at least 100 years past the initial spin-up period. After about year 100, the model continues to warm but at a much slower rate. This latter period (years 101-500) of fairly constant, more moderate drift is used as the trend analysis period for the CM2.0 runs, except as described in the text.

Figure 1 (a) indicates the three 140-year periods of the CM2.0 control run that correspond to the three 140-year all-forcing historical runs for that model (h1-h3). Note that the first CM2.0 historical run (h1) was initiated from a point in the control run (year 1) in which the model was continuing to drift warm at a substantial rate. In retrospect, a better choice of initial condition for h1 would have been after year 100 of the control run. However, the need to have several completed historical scenario runs for an IPCC model intercomparison project necessitated the choices made. To adjust for the effects of control run drift, the 140-year time series from historical run h1 are adjusted by subtracting the trend of the control run over the same 140-year time interval. Historical runs h2 and h3 are adjusted by subtracting the trend of the control run over the longer subsequent period (years 101-500).

The evolution of global mean T_{ref} in the CM2.1 control run is shown in Fig. 1 b. Following a pre-control run spin-up period of 220 years (of which only the last 100 years are shown), the model exhibits a moderate warming trajectory of about $0.1^{\circ}\text{C } 100 \text{ yr}^{-1}$ for years 1-300 of the control run, followed by a more gradual warming trend during years 301-900. For our trends analysis in this study, we consider these two epochs (years 1-300 and 301-900) separately, and remove the long-term drift from them separately. Note that all five all-forcing historical runs for CM2.1 (H1-H5) were run during the period in which the control run had a moderate but relatively stable degree of drift. These historical runs are adjusted for control run drift by subtracting the trend of the control run over years 1-300.

The red curves in Fig. 1 (a and b) are the observed global-mean annual-mean temperature anomaly series based on the HadCRUT2v data set. An arbitrary vertical offset has been added to the anomaly series to display it for reference alongside each control run annual mean temperature series. The pronounced warming in the observed global mean series clearly exceeds the internal variability of either the CM2.0 or the CM2.1 control runs. This is evident in Fig. 1 after accounting for the long-term adjustments mentioned above, by comparing observations with the later (relatively low-drift) centuries of the control runs. This finding is consistent with numerous previous studies (e.g., Stouffer et al. 2000) which find that the observed global mean surface temperature warming exceeds changes due to internal climate variability as simulated by long control runs of several global coupled climate models.

b. Geographical distribution of interannual variability

A preliminary assessment of a climate model's internal variability can be made by comparing maps of the local interannual standard deviation of surface temperature between model and observations. The standard deviation of annual means includes variance on time scales from two years out to the length of the record, and thus includes substantial contributions from time scales that are much shorter than the multidecadal trends that are the main focus of our study. Nonetheless, the comparison of interannual standard deviations provides a useful benchmark before focusing on longer time scales.

Figure 2 shows the interannual standard deviations for observations (a), CM2.0 (b), and CM2.1 (c). The model fields are based on SST (ocean regions) and T_{ref} (land regions) for years 101-200 for versions of the CM2.0 and CM2.1 control runs with constant 1990 climate forcings. All observed and modeled series in Fig. 2 have been detrended using linear least squares regression. The observed field is based on years 1949-2003 from the combined SST/land surface temperature HadCRUT2v dataset. While there is a similar overall pattern of the standard deviation fields for the models and observations, with enhanced variability over continental regions relative to the surrounding oceans, and locally enhanced SST variability in the tropical Pacific associated with El Niño, a clear deficiency of the models is the greater than observed simulated interannual variability over many land regions. For CM2.1, the El Niño region variability is also excessive, and in both models the most pronounced El Niño variability is displaced westward from the observed maximum location near the South American coast (see Wittenberg et al. 2005 for further details).

The enhanced variability over land regions in high latitudes does not appear to be attributable to excessive El Niño variability, as seen by comparing the maps for CM2.0 (b) and CM2.1 (c). Despite less excessive El Niño variability compared to CM2.1, CM2.0 shows clearly excessive variability over continents in high latitudes. Since CM2.0 and CM2.1 also have a number of parameter differences which could potentially affect high latitude continental variability, we also show Fig. 2 d, which is a version of CM2.1 with reduced El Niño region variability. This sensitivity experiment (d) was based on previous modeling experience (e.g., Wittenberg et al. 2005) that the model's El Niño variability is sensitive to the cumulus momentum transport in the atmospheric model. For the sensitivity experiment in (d), cumulus momentum transport was turned off, producing a version of CM2.1 with weaker El Niño variability. The modified model shows a reduction, compared to CM2.1 (c), in the excessive interannual variability biases in a number of tropical and subtropical land regions, including northern South America and the southern U.S. However, substantial biases remain in the northern extratropics, Australia, and southern South America, confirming that the extratropical biases are not primarily due to excessive El Niño variability in the model. The cause of the excessive simulated variability in the models is a topic of continuing investigation.

An important implication of the results in Fig. 2, in the context of the present study, is that there is no evidence that the CM2.0 and CM2.1 models substantially underestimate local internal climate variability, at least as aggregated over all resolvable time scales in annual mean data. Rather, the models show a tendency to overestimate this variability. This suggests that the models may provide a conservatively high estimate of internal climate variability when assessing whether observed multi-decadal trends are within the range of expected internal climate variability. We will return to this issue in a later section, when we assess the variance spectra and standard deviation of observed and simulated global mean temperature in the historical simulations.

5. Historical Simulations: Global Mean Temperature

In this section and section 6, we analyze the CM2.0 and CM2.1 historical simulations, beginning in this section with an analysis and comparison of global mean surface temperature time series and their variance spectra.

a. Simulated vs. observed global mean temperature time series

Global mean surface temperature anomaly time series for various historical forcing runs are compared with observed temperature anomalies in Fig. 3. For these series, the model data have been masked out for periods and locations where data is missing according to the HadCRUT2v observed data set. Sea surface temperature is used for model grid points with more than 50% ocean coverage, while T_{ref} is used for all other model grid points. Prior to computing the global means, the model time series at each grid point have been adjusted to account for long-term drifts in the control simulation as described in section 4. The data are displayed as annual mean anomalies relative to the 40-year means for 1881–1920. The thick black curve in the panels denote observations (HadCRUT2v) while the thick red lines are the ensemble means of various historical runs. Thin dashed green lines are the individual ensemble members.

The time series for both the CM2.0 (a) and CM2.1 (b) all-forcing experiments are in good agreement with the observed series in terms of the overall warming through the 20th century. There is a clear tendency for the observed anomalies to fall within the range of the model ensemble members, with some notable exceptions as discussed below. In general agreement with observations, both all-forcing ensembles show a tendency for periods of fairly rapid warming early in the 20th century, followed by a few decades of relatively little warming near the mid-20th century, with resumed rapid warming from about the mid-1960s (mid-70s in the observations) through the end of the experiments in 2000.

The agreement between the CM2.1 anthropogenic-only series (e) and observations is also quite good, with the largest discrepancy shown being the strong temporary warming period around 1940 in the observations, which is not reproduced in the model runs. In contrast, the CM2.1 natural-only forcing experiments (d) clearly fail to reproduce the strong warming in the late 20th century seen in the observations, although they appear to contribute significantly to the early 20th century warming.

Other notable features of the global mean temperature curves for the all-forcings and natural-only forcing runs are the pronounced, short-lived cooling periods clearly evident in the model results in the 1880s, 1900s, 1960s, 1980s, and 1990s. These correspond to periods following strong volcanic aerosol forcing associated with the Krakatau (1883), Santa Maria (1902), Agung (1963), El Chichón (1982), and Pinatubo (1991) eruptions, respectively. In the all-forcing runs (a, b), the period of resumed rapid global warming in the late 20th century begins following the cooling associated with Agung in the mid-1960s.

There is an impression from the all-forcing results (a, b) that the model's cooling response to several of the volcanic eruptions is larger than observed, particularly for Krakatau, El Chichón, and Pinatubo. However, this initial impression requires further investigation, since the timing of short-lived climate fluctuations associated with El Niño in the model and in the real world can complicate such interpretations (Santer et al. 2001; Soden et al. 2002). For example, with regard to Pinatubo, Fig. 4 of Soden et al. indicates that without the warming influence of El Niño, lower tropospheric temperatures during 1992

would have been roughly 0.1°C cooler on average than what actually occurred. This compares with the model error (model vs observations) of about 0.25°C for 1992 in CM2.1 (Fig. 3 b). In addition, ensemble mean tropical Pacific SSTs in our model simulations were below normal in 1992 for both CM2.0 and CM2.1 (not shown), presumably by coincidence¹. Similarly, for the El Chichón eruption, the occurrence in the same year of the strong 1982-83 El Niño event probably counteracted some of the global cooling from the eruption, while in the CM2.0 and CM2.1 ensemble means, equatorial Pacific temperatures were unusually cool, not warm, during this period. Thus, the model/observational discrepancies for Pinatubo and El Chichón appear to be partially attributable to El Niño influences, although further investigation (outside the scope of this paper) is needed.

The large cooling discrepancy in the 1880s is fairly prolonged in both models, covering the latter two-thirds of the decade (Fig. 3 a, b). In contrast to the late 20th century, these discrepancies do not appear to be exacerbated by strong El Niño influences. A likely strong contributor to the relatively prolonged cooling behavior is the occurrence of multiple significant volcanic eruptions during the period (Krakatau in 1883, Tarawera in 1886 and Bandai in 1888). The discrepancy between the models and observations during this period is unresolved, but could be due to problems in the observations (eruption or surface temperature data), radiative forcing specification, or the model's sensitivity to short time-scale volcanic forcing. Tree-ring data, used as a proxy for Northern Hemisphere summer temperatures, provide some independent support for the observed temperature records indicating little global cooling during the period (Jones et al. 1995). Further investigation of individual forcing factors and internal variability will be needed for more definitive conclusions.

The low simulated temperatures during the 1880s in CM2.0 and CM2.1 natural-only and all-forcing runs could significantly affect trends that begin in the late 1800s. Later in this report, we present some long-term trend analyses which use 1901 as the starting year -- a choice of starting year that reduces the influence of these Krakatau-era discrepancies on the trend statistics.

In a study with the GFDL R30 coupled model, Broccoli et al. (2003) concluded that the simulated response of that model to volcanic eruptions since the late 1800s was larger than observed, based on a superposed epoch analysis. We note that their simulations used a different prescribed volcanic forcing derived from another global model by Andronova et al. (1999) and was a more idealized implementation of volcanic forcing than used here.

The rapid warming in the early 20th century appears likely due to a combination of anthropogenic and natural forcings (panel b) including increased greenhouse gases, reduced negative volcanic forcing following Krakatau and other eruptions just discussed, and a positive contribution from solar variations. This finding is in agreement with earlier studies (e.g., Stott et al. 2000). The results in panels (d and e) also suggest that either natural-only forcing or anthropogenic-only forcing, in combination with unusually strong temporary warming from internal climate variability, can reproduce an early 20th century warming in the model that is fairly close to the observations. For example, one anthropogenic-only ensemble member shows a temporary warming peak in the mid-1940s, similar to the observed peak a few years earlier. An even closer “anthropogenic forcing plus internal variability” analog for the early 20th century global warming was shown for an earlier model

¹ Adams et al. (2003) present empirical evidence, based on volcanic and paleoclimate proxies, for a link between explosive volcanic eruptions and the probability of El Niño occurrence the following winter. The limited ensembles in our experiments do not provide model-based evidence for such behavior.

by Delworth and Knutson (2000). However regarding the late 20th century warming, the results in Fig. 3 have little ambiguity: the rapid warming during the late 20th century in CM2.1 is a consequence of anthropogenic forcing, since the runs with natural-only forcing fail to reproduce this strong observed late 20th century warming. Solar variability contributes little positive radiative forcing during the late 20th century while volcanic eruptions produce strong temporary negative forcings late in the century—particularly from Pinatubo (Ramaswamy et al., in preparation).

The remaining panel in Fig. 3 (panel c) examines the linearity of the model's response to the forcings by comparing the anomaly ensemble mean series from the CM2.1 all-forcings ensemble to the sum of the ensemble mean anomalies from the CM2.1 natural-only and anthropogenic-only forcing runs. The curves are very similar, indicating that the model's global response to these forcings is approximately linear. A similar result has been found in previous studies (e.g., Ramaswamy and Chen, 1997; Haywood et al. 1997; Gillett et al. 2004; see also Hansen et al. 1997 and Ramaswamy et al. 2001).

b. Variance spectra of global mean temperature

Figure 4 shows variance spectra of the observed global mean temperature series (1871-2000, dark red curve in each panel) in comparison to variance spectra of 130-year segments (1871-2000) from the a) CM2.0 and b) CM2.1 all-forcing historical runs. The ensemble spectra from the all-forcing runs (medium black lines) indicate that the model's variance spectrum is fairly similar to observations in terms of both its general shape. However, there is a clear tendency toward excessive variability, compared to observations, on almost all time scales, especially in CM2.1. In fact, the CM2.1 ensemble spectrum lies above the observed spectrum at all frequencies shown except the lowest frequency (~65 yr). The CM2.0 ensemble spectrum also tends to lie above the observed spectrum, although not as consistently as for CM2.1.

The standard deviations of the raw and detrended annual-mean and the detrended 10-yr running mean global-mean time series for observations and the all-forcing historical runs are presented in Table 1. The annual statistics aggregate variability across all time scales resolvable by annual means in the 130-yr records, rather than examining different frequency bands of the spectrum. For detrended data annual, the average standard deviation of the CM2.1 all-forcing runs exceeds the observed standard deviation by about 38%. The CM2.0 all-forcing scenario standard deviations also substantially exceed the observed value (by about 25%). The fact that all eight independent CM2.0 or CM2.1 all-forcing runs have larger detrended standard deviations than the detrended observations indicates that model's global annual-mean variability is very likely larger than observed variability. For example, if the modeled and observed standard deviation samples came from the same population, the probability that the observed standard deviation be lower than any of the eight model standard deviations, as in Table 1, would only be about 11%. Standard deviations of 10-yr running mean (detrended) data, which focus on low-frequency variations, show a smaller positive bias of variability in the models compared with observations (+7% and +11% for CM2.0 and CM2.1, respectively).

As already discussed in the context of the control simulations, the excessive global mean surface temperature variability in the models, likely related to the excessive El Niño variability in CM2.1 as well as the excessive interannual variability over continental land surfaces, has the consequence that the models are likely to be providing a high (conservative)

estimate of the potential contribution of internal climate variability to observed long-term trends, as assessed in the following section. On the other hand, the excessive variability in the models will increase the spread among ensemble members, making it more likely that observations will fall within the spread of the individual ensemble members.

6. Assessment of Regional Surface Temperature Trends

In this section, the surface temperature trends in the models and observations are compared for different regions. We focus first on the latter half of the 20th century (1949-2000), a period during which global data coverage was substantially more complete than during the late 19th and early 20th centuries. Later in the section we attempt some trend analyses on the longer (century) scale in regions with at least moderate data coverage. Finally the temporal character of low-frequency temperature variations in various “case study” regions are examined using area-averaged time series.

a. 1949-2000 trends

Regional surface temperature trends in the observations and models are assessed for the period 1949-2000 by comparing the observed pattern of trends with the internally generated 52-yr trends in the control runs and with the trend patterns from the 1861-2000 historical simulations over the same 52-yr period. The control run sample segments and model segments from the historical runs are masked with the time-varying observed missing data mask prior to computing the trend samples. Areas with less than 80% data coverage during the period are not included in the analysis and appear as white regions in the maps.

An assessment of the annual mean trend maps obtained by combining the CM2.0 and CM2.1 ensembles is shown in Fig. 5. The observed trend map (a) shows pronounced warming ($2\text{--}4^{\circ}\text{C } 100 \text{ yr}^{-1}$) over many northern middle- and high-latitude continental regions with a broad region of more gradual warming ($1\text{--}2^{\circ}\text{C } 100 \text{ yr}^{-1}$) over much of the Indian Ocean/western Pacific warm pool region, Australia, the eastern tropical Pacific, and South Atlantic (north of 40°S). Regions of cooling, smaller in overall extent than the warming regions, are seen in the north Pacific, much of the North Atlantic basin, the southeastern U.S., New Zealand and vicinity, and a few other small regions.

The observed trend assessment, based on the model control runs (b, color shaded regions) indicates that the observed warming trends over much of the globe are unusual in comparison to the control run. About 69% of the global area examined has observed trends which are outside the 5th-95th percentile range of the trends for that region from the combined control run samples (23-nonoverlapping 52-yr segments). The areas of significant trends are dominated by warming trends, with only small regions of the globe identified as having significant cooling trends.

The ensemble mean trend map for the eight CM2.0 ($n=3$) and CM2.1 ($n=5$) all-forcing runs in Fig. 5 (c) shows a broad-scale warming pattern over much of the globe with enhanced warming over many continental regions. A large area of cooling is simulated in the north Pacific similar to the observations although extending further north and east in the basin and not as far equatorward as the observed cooling trend region. Smaller secondary regions of cooling appear in extreme eastern Asia, central Asia, and the North Atlantic. The all-forcing simulations do not show particularly strong warming trends in middle and high latitude land regions of the northern hemisphere, in contrast to the observations.

Figure 5 (d) shows an assessment of where the ensemble-mean simulated trends and observed trends are significantly different from one another. The field shown is a map of model-minus-observed trend differences (i.e., c minus a), but with a modified local t-test applied such that areas are blacked out where the null hypothesis that the simulated and observed trends are equal cannot be rejected at the 0.10 level using a two-sided test. By this definition, the remaining color-shaded regions are areas where the simulations are significantly different from the observations. The modified t-test used (following Knutson et al. 1999) is analogous to a standard two-sided two-sample t-test except that the long control runs are used to estimate the variance of the trends (assumed to be the same for both the model and observations). The n_1 and n_2 parameters for the t-test are $n_1=1$ for the observation and $n_2=8$ for the eight-member CM2.0/CM2.1 ensemble. The modified t-test statistic assumes 23 degrees of freedom based on the total number of non-overlapping 52-yr segments from the control runs (assumed independent) that are used to estimate the variance of 52-yr trends ($n=23$).

The t-test results for the all-forcing runs in Fig. 5 (d) indicate that these runs have significantly less warming than observed during 1949-2000 in a number of regions, including much of northern Asia, and parts of Canada, the Indian Ocean/warm pool region, the southern Indian Ocean, and the South Atlantic. Most of the regions of inconsistency are areas where the all-forcing runs do not warm rapidly enough. Some smaller regions where the model warms too rapidly include parts of the central tropical Pacific, northern tropical Atlantic, South America and equatorial Africa. The tests indicate that the all-forcing ensemble is not significantly different from the observations (black regions) in the North Pacific cooling region and most of the North Atlantic cooling regions in panel (a). The percent of global area tested where the ensemble simulation has significantly different trends from observations is about 33%, as compared with 69% for the control run comparison in (b).

The ensemble mean simulated trend for the ($n=4$) natural-forcing runs (Fig. 5 e) shows a cooling in most regions during 1949-2000, and thus the disagreement with observed trends is quite large. About 70% of the global area examined has significantly different simulated trends than the observed for this period (panel f). This is similar to the degree of inconsistency in the control run (no external forcing) comparison.

The ensemble mean results for the anthropogenic-only runs (Fig. 5 g) show greater warming rates in many regions than the all-forcing runs (c). In terms of cooling regions, the ensemble mean map shows similar regional features to the all-forcing runs (c) in the extratropical North Pacific and somewhat more cooling in the high latitude North Atlantic. The t-test assessment (panel h) shows a slightly smaller percent area with significant differences (29%) than the all-forcing ensemble (33%), with more of a balance between areas with negative and positive trend differences.

Although not shown here, we have performed assessments for the CM2.0 and CM2.1 ensembles separately and obtain broadly similar results for the two sets of ensembles. This provides some justification for our combining the results from the two models into a single combined assessment in Fig. 5.

Seasonal versions of the combined CM2.0/CM2.1 assessments for the observed trend and the all-forcing ensemble are shown in Fig. 6. Broadly similar results were obtained for the CM2.0 and CM2.1 ensembles separately (not shown). Many of the characteristic features of the annual mean results just discussed also appear in the seasonal mean assessment results.

A distinctive feature of the observed seasonal trend maps for DJF and MAM is the very pronounced warming in extratropical northern hemisphere land regions (e.g., upper left panel of (a) and (b)). This seasonally modulated character of the warming is largely absent in the simulated ensemble-mean trend maps. In addition, the area with significant differences between model and observations is much larger in the winter and spring than in summer and fall in those regions. The northern Pacific cooling feature appears in all seasons in the observations and the model ensembles (both in CM2.1 and CM2.0 individually as well, though not shown). The high latitude North Atlantic cooling feature appears in all seasons in the observations, but much more weakly in the model ensemble. However, the significance test results indicate that the observed cooling trends over the North Pacific and North Atlantic are mostly not significantly different from the all-forcing model ensemble trends.

b. 1901-2000 trends

Trend maps for the entire 20th century (1901-2000) are shown for the observations and model historical runs in Fig. 7. The more sparse data coverage in the first half of the 20th century presents potential problems for such an analysis. Therefore, we use two different percent data availability criteria to screen the observed results in order to highlight regions with poor vs moderate data coverage. In the middle column (a,d,g,j,m) the observed and ensemble mean historical run trends are displayed for all grid points with at least 5% temporal coverage during each of five 20-year periods (1901-1920, 1921-1940, etc.). Effectively, these screens apply to the first half of the 20th century, since coverage generally has been much greater since around 1950. In the right column, a more stringent criterion (30%) is used. For the model results in the middle and right columns, the model data has been masked with the time-evolving observational mask. In the left column, model trends are shown for all grid points and with no data masking.

Some broad-scale differences are immediately apparent between the observed and simulated all-forcing ensemble-mean 100-yr trends (e.g., a vs. d or g). For example, the observations show a tendency for larger warming trends in the extratropics than in the tropics, in contrast to the all-forcing ensembles (e.g., d, g), where the simulated warming in the tropics is at least comparable to that in the extratropics (for regions with sufficient observational data coverage). The relatively strong observed warming trends over the northern continental regions, parts of the South Atlantic, and extratropical southern Indian Ocean are more pronounced than any large-scale trends in the models. Another clear difference is the presence of a large cooling region in the extratropical North Pacific in the historical simulations – a feature which appears to be absent in the observations. Some of the local simulated vs observed differences noted above are undoubtedly attributable to internal climate variability; however, we have not yet performed t-tests on the 100-yr trend comparisons due to the need for a longer control integration than is presently available.

Comparing the left column of Fig. 7 with the masked model results in the center and right columns gives some indication (by inference) of the possible effect of missing observations on features in the observed trend fields (a, b). We sub-sample the model data using only times where the observations are available, and compare with the “perfect” case of no missing data. Thus, any differences in regional details between panels i and j are due to the deletion of missing data periods from the model data in regions where temporal data coverage is sufficient to meet the minimum availability criterion. Comparison of the masked and unmasked model trend maps suggests that while the missing data can affect detailed

features of the trend maps, its effect on most large-scale features does not appear to be very critical, even over the Pacific sector where coverage is relatively sparse.

The 100-yr trends over much of the North Atlantic Ocean tend to be in fairly good agreement between models and observations with both showing mostly moderate warming trends, with somewhat enhanced warming in the tropical North Atlantic, and a region of cooling trends in the high latitude North Atlantic. However, in the southeastern U.S.--another region of localized cooling in the observations--the simulations do not show a cooling trend, a finding that will be examined further in the regional case studies analysis below.

The 100-yr trend maps for the CM2.1 natural-only forcings ensemble (Fig. 7 i-k) show mostly cooling trends, in poor agreement with observations. The CM2.1 anthropogenic-only forcing ensemble (l-n) is in better agreement with observations than the natural-forcing runs, and perhaps even than the all-forcing runs, although effects of internal climate variability have not been formally assessed. In particular, the anthropogenic-only ensemble trends show more evidence of enhanced warming over high-latitude continental regions than the CM2.1 all-forcing ensemble (f-h).

While Fig. 7 shows interesting features for the observations and clear differences with the historical simulations, a caveat is the relatively sparse data coverage and other data issues as one extends such analyses further back in time. These issues, including uncertainties due to adjustments to SST observations prior to the early 1940s (e.g., Folland et al. 2001) imply a greater uncertainty in the observed trend details for Fig. 7 than for Fig. 5. Although not used in the present study, an alternative approach would be to use currently available SST historical reconstructions (e.g., Kaplan et al. 1998; Rayner et al. 2003; Smith and Reynolds 2003) which attempt to address at least the missing data shortcoming by filling-in data gaps in the temperature records based on space-time statistical methods.

c. Regional Case Studies

The trend maps in the previous discussions are useful for revealing coherent spatial patterns and other features in the trend fields, but they require the rather arbitrary selection of starting and end dates, to which linear trend results are sensitive. An alternative to trend maps are direct comparisons of observed and modeled time series. Due to space considerations, only a limited number of “case study” regions are chosen for focused examination in this study. Each of the time series has been adjusted for control run drift, screened with the observed data coverage for the region, and low-pass filtered (10-year running mean) to emphasize low-frequency variations, which are most relevant to the present study. The time series are anomalies with respect to years 1881-1920 for these figures. The areas used to construct the regional time series are identified in Fig. 8. We also attempt to show the possible effect of missing observations on the observed regional time series by statistically comparing masked versus un-masked versions of the model time series.

1) GLOBAL, EXTRATROPICAL, AND TROPICAL MEANS

Figure 9 (a, b) shows the observed global mean temperature versus the all-forcing historical run ensemble members for CM2.0 (a) and CM2.1 (b). Figure 9 (c) shows the ensemble means from the all-forcing (n=5), natural-only forcing (n=3), and anthropogenic-only forcing (n=3) CM2.1 runs. The simulated and observed global mean curves (a-c) are in fairly good agreement for the all-forcings and anthropogenic-only forcing runs, but not for

the natural-only forcing runs, as discussed previously for Fig. 3. The global mean results are provided here for reference comparison for the regional case study time series.

Panel (c) of Fig. 9 also shows the estimated uncertainty due to missing observations on the observed global mean temperature (10-yr means). The shaded region in (c) is the ± 2 standard error range about the observations, which was estimated by differencing model 10-yr running mean series obtained with and without use of the observational mask. The error estimate is based on the sample of eight available all-forcing scenario runs (three for CM2.0 and five for CM2.1) and eight natural-only (4) or anthropogenic-only (4) forcing runs. The error range decreases over time from about 0.12°C in the late 1800s to about 0.06°C in the late 20th century. Note that this is only a partial error estimate, as it does not attempt to show the uncertainty range due to errors in the available data or the uncertainty range due to internal climate variability (i.e., the range of results in a hypothetical world with an “ensemble of observations” available). For example, Fig. 2.8 of Folland et al (2001) includes uncertainties due to bias corrections in SSTs prior to the early 1940s and due to urbanization, both of which appear to have a substantial impact on the uncertainty estimates. However, since it is straightforward for us to estimate at least the missing data component of the uncertainty for the global mean series as well as the regional series, these partial error estimates are included in our case study figures.

For the northern extratropics (d-f) the all-forcing simulations show a similar time evolution to the observed in some respects, including relatively rapid warming early in the 20th century, relatively little warming or even some cooling from the 1930s (1940s in the observations) to the 1970s, and resumed strong warming from the 1970s onward. The southern extratropics, shows a more monotonic warming behavior in the all-forcing simulations, in agreement with observations. There is some tendency for the all-forcing simulations to warm less than the observations in the extratropics, as the observed curve tends to lie on the upper edge or slightly above the range of the model ensemble members. In the tropics (j-l), the all-forcing runs tend to warm slightly more than the observations, with the observed curve typically falling on the lower edge or below the range of the model ensemble members. The natural-only ensemble again shows a clear deficiency in not simulating the strong late-20th century warming in any of the regions in Fig. 9. The strong cooling in the 1880s discussed earlier appears most pronounced in the model in the northern extratropics (d-f) with a lesser expression in the tropics (j-l) and only a minor expression of the cooling in the southern extratropics (g-i).

The finding that the all-forcing model tends to warm too rapidly in the tropics and not rapidly enough in the extratropics implies that the model simulates less poleward amplification of warming than observed during the 20th century. To investigate the poleward amplification issue in more detail, we have examined time vs latitude plots (not shown) of zonal mean temperature anomalies for the observations, historical runs, and a future climate change scenario (CM2.1 IPCC SRES Scenario A1B to 2100). While the historical run ensemble mean shows less poleward amplification of warming than the observations, the A1B future scenario shows the northern hemisphere extratropics warming more than the tropics in the model, while the deep southern hemisphere warms very little. A similar pattern of response to the A1B scenario is seen in $+1\%/yr$ CO_2 transient experiments with CM2.0 and CM2.1 (Stouffer et al., submitted for publication). Doubled CO_2 equilibrium experiments, with the atmospheric components of CM2.0 and CM2.1 coupled to a slab ocean model, show significant poleward amplification of warming in both hemispheres

(Stouffer et al., submitted for publication). These future climate scenario results, showing marked polar amplification of warming in large warming perturbation experiments, suggest that internal climate variability could be playing a role in the apparent discrepancies in the northern hemisphere in the historical runs. Further investigation, outside the scope of this paper, will be needed to clarify this issue.

The impact of missing data on the observed curves in Fig. 9 is estimated to be greatest for the southern hemisphere extratropics (i). The error range is also relatively large (about ± 0.15) in the 1880s in the northern extratropics.

Finally, we note that much of the low-frequency (multi-decadal) fluctuations in the all-forcing runs appear to have a broadly similar timing or phasing across the different ensemble members, as well as for the observations. For example in the northern extratropics (d, e), the temporary period of slight cooling from around the 1950s to around 1970, followed by strong warming, is seen in the envelope of the ensemble members. However, the ensemble mean for CM2.1 all-forcing (f) also suggests a longer break between major warming periods than seen in the observations. A second example is the pronounced cooling episode seen in the 1880s in all panels. This common behavior among ensemble members indicate that these features are being forced in the model, and perhaps in observations, by volcanic activity and other specified climate forcing agents, as opposed to being internal climate variations.

2) TROPICAL REGIONAL INDICES

Various tropical and subtropical regional indices are shown in Fig. 10. The Indian Ocean/western Pacific warm pool index (a-c) shows a relatively pronounced warming trend in the observations. The observed curve shows fairly rapid warming during the 1920s and 1930s, followed by cooling during the 1940s, and then a relatively monotonic warming trend beginning around 1950. The strong monotonic warming since 1950 in this region produced a broad-scale statistically significant warming signal in an earlier model-based assessment (Knutson et al. 1999; see also Hoerling et al. 2004)—a finding which was reconfirmed in the present study as well (Fig. 5 b). A similar, though slightly noisier, behavior was evident for the tropics as a whole (Fig. 9 j-l). The simulated warming trend over the 20th century is more monotonic for the ensemble mean of the CM2.1 all-forcing runs (Fig. 10 c) than the observations, although individual ensemble members (a, b) show a few multidecadal variability episodes similar in amplitude the observed fluctuation from 1920-1950. The timing of the model-generated events does not coincide with the observed one. The models tend to warm more rapidly than observed prior to 1950, although from 1950 to 2000 the observed warming has tended to be more rapid than the simulated warmings. Thus the observed warming nearly catches up with the simulated warming by the end of the period, as the observations reach the lower edge of the ensembles of the all-forcing runs for both CM2.0 and CM2.1.

In the tropical East Pacific, the simulated and observed time series are characterized by a much greater degree of noise relative to the warming signal, due to the influence of El Niño and perhaps Pacific decadal variability. Despite the high noise level in this region, the observed 10-yr running mean temperatures since about 1980 are the warmest in the record. During a temporary warming around 1900, temperatures reached levels within a few tenths of a degree of those of the most recent decades. The all-forcing ensemble members, particularly for CM2.1, tend to show a more monotonic warming evolution in this region,

which results in a period of several decades, from about 1940 to 1980, when the model appears substantially warmer than the observations. However, from about 1980 on, the strong late century warming in the observations has considerably narrowed the discrepancy with the simulations such that the recent observations are again within the envelope of the ensembles. This is similar to the behavior for the Indian Ocean/western Pacific warm pool (a-c) discussed above.

Although the estimated errors in the observations due to missing data are larger as one goes further back in the records, the error is estimated to be smaller than for the southern hemisphere extratropics, or even the global mean, as seen by comparing the shaded regions in Fig. 10 (c, f) with Fig. 9.

The tropical North Atlantic region from 10-20°N was selected for particular focus, as this region is known as the main development region for Atlantic hurricanes that evolve from tropical easterly waves emanating from Africa (e.g., Goldenberg et al. 2001). The discrepancies noted above for the Indian Ocean and eastern tropical Pacific simulations for the 1940s through 1980 are not evident in the tropical North Atlantic index Fig. 10 (g-i). In fact the model ensembles appear to capture many aspects of the temporal evolution of the observed index for this region, including a tendency for more warming in the first half of the 20th century than during the second half of the 20th century. The model tends to be too cool during the 1880s, similar to the behavior discussed for the global mean and other regions, and slightly too warm at times during the early 20th century.

The extension of the observed index using data through 2004 (g, h) reveals a continued warming in the region which has now brought the 10-year running mean index to unprecedented warm levels, a finding which also appears consistent with the upward trend in all the all-forcing model runs during the 1990s. This region is not one identified as having significant warming over the past half century in Fig. 5. However, the time series results indicate that this may have been a result of strong multidecadal variability (either forced or internal) obscuring the century-scale warming trend signal over the past half century. If this interpretation is correct, the warming late in the 20th century in this region represents the emergence of a long-term warming signal from the background of substantial multidecadal variability.

Fig. 10 (j-l) shows temperature anomalies for a large region of the Atlantic that extends from 10°N to 45°N, thus encompassing most of the subtropical North Atlantic. (Note that this region, extending equatorward to 10°N, overlaps the tropical North Atlantic region just discussed.) Relatively good agreement between all-forcing simulations and observations is also seen for this larger region, particularly for CM2.1. As was the case for the tropical North Atlantic, this region has been characterized by relatively little net warming over at least the last half century, but may well be entering a period of pronounced warming. The underlying warming trend is likely driven by anthropogenic forcing (as inferred from the CM2.0 and CM2.1 simulations) as natural-only forcing runs fail to simulate the strong late 20th century warming. The multi-decadal variations are conceivably related to fluctuations of the North Atlantic thermohaline circulation as inferred from the NAO-related heat and wind stress forcing for the 20th century (e.g. Fig. 5 in Eden and Jung, 2001). The timing of fluctuations in these Atlantic regions appears to be similar to that of the northern extratropics as a whole (Fig. 9 d-f), and a muted form of these fluctuations is even discernible in the global mean (Fig. 9 a-c).

3) EXTRATROPICAL REGIONAL INDICES

In Fig. 11 are time series comparisons for several extratropical regional indices. The high latitude North Atlantic region (45-65°N) was chosen for a case study because of the intriguing cooling trends shown there in many of the modeled and observed trend maps for both the last half century and the full 20th century. The observed 10-yr running mean time series for this region (a-c) shows pronounced multidecadal variability with particularly strong cooling period from the 1950s to the 1970s and rapid warming in the 1990s. The observed curve typically falls within the range of the all-forcing ensemble members, particularly for the larger (n=5) CM2.1 set of ensembles. During the second half of the 20th century, several of the CM2.0 and CM2.1 individual all-forcing simulations (and the CM2.1 all-forcing ensemble mean) tend to follow a similar evolution to the observations in the late 20th century. This agreement may be coincidental, as a few ensemble members show a quite different temporal evolution.

A notable characteristic of the model time series for this region is the large scatter between ensemble members (note the increase in scale of the vertical axis compared with some of the previous figures), indicating lower signal to noise for radiative forcing response in this region due to the presence strong internal climate variability. The impact of missing observations is estimated to be moderate (Fig. 11 c) during the period prior to 1920, but is smaller than the scatter of the different ensemble members.

The extratropical north Pacific index (Fig. 11 d-f), like the high latitude North Atlantic index, has large multi-decadal variations in the observations and model simulations, and relatively large degree of scatter between individual all-forcing ensemble members. The observed cooling trend in this region during the last half of the 20th century, a pronounced feature of Fig. 5 (a), appears to be part of a multidecadal variation, since Fig. 11 (f) shows a period of strong warming just prior to 1950. The observations typically fall within the range of the model ensemble members, although the CM2.1 all-forcing ensemble mean time series is not that similar to the observed series, apparently reflecting the impact of the large internal climate variability in the region. The potential impact of missing data appears to be substantial in this region, particularly prior to 1920. The error range is of the same order of magnitude as the model's internal variability (scatter between ensemble members).

Northern Asia (40-70°N) was chosen for a case study (Fig. 11 g-i) because the models did not warm as strongly as observed in this region according to the trend map analyses (e.g., Figs. 5-7). The observed time series for this region shows fairly monotonic warming, with a relatively moderate warming trend from the late 1800s through about the 1970s, and a much more rapid warming in recent decades. The total warming since the late 1800s in this region appears to exceed 1.5°C, or about twice the global mean warming. Such a strong warming is not apparent in any of the all-forcing model ensembles, although some individual ensemble members show some evidence for large internal climate variability in this region. In that regard, a caveat is the tendency in the model for excessive simulated interannual variability in this region (Fig. 2), which suggests that the internal climate variability may be exaggerated in this region even on decadal time scales. In fact, the observed time series gives the appearance of only fairly modest decadal-scale variability, with the dominant feature being either very low-frequency (i.e., multi-century) variability or a long-term monotonic warming trend. We will return to this region for additional analysis in the next section.

The southeastern U.S. is another area of striking disagreement between the models and observations (Fig. 11 j-l). The observations show a very gradual warming trend of about

0.25°C from the late 1800s to the mid-20th century, following by a pronounced cooling into the 1960s and 70s, followed by a partial warming recovery since around 1980. In contrast, the CM2.1 all-forcing ensemble and its individual ensemble members show a consistent gradual warming through the entire 130-yr period. CM2.0 shows some tendency for long-term warming in two of three all-forcing ensemble members, but the third shows little warming over the century. Thus the warming appears to be small relative to internal climate variability in CM2.0. A caveat is that this region, like northern Asia, is one in which the model's simulated interannual variability appears excessive (Fig. 2) which suggests the possibility of excessive simulated multi-decadal climate variability in the region. The time series for the observations in the southeastern U.S. shows relatively small levels of decadal variability on shorter decadal resolved time scales, but contains a single very low frequency fluctuation during the second half of the 20th century. The unusual behavior in the observed temperatures in this region and its contrast to the model simulations is a topic worthy of further investigation. Missing data effects appear to be minimal (Fig. 11 l).

7. Analysis of Arctic Oscillation Impacts

A notable regional discrepancy in the simulations versus observations is the lack of pronounced warming over northern Asia in winter and spring late in the 20th century (e.g., Fig. 6, Fig. 11 g-i). Surface temperatures, including a substantial fraction of recent winter warming trends, in this region are known to be associated with (i.e., linearly congruent with) the North Atlantic Oscillation (NAO, Hurrell 1995) and Arctic Oscillation (AO) (e.g., Thompson and Wallace 2000; Thompson et al. 2000). Therefore, a question arises as to whether the temperature discrepancies can be explained in terms of discrepancies in the observed and simulated AO behavior. To investigate this issue, we have performed additional analyses on northern Asia time series from the CM2.1 all-forcing runs, where we estimate the potential influence of recent observed positive AO anomalies on simulated northern Asia temperatures.

Fig. 12 (c) shows the observed AO index through 2004 (10-yr running means). The index shows little trend over most of the 20th century, followed by a large shift toward positive (high index) values from around 1980 to the early 1990s, and a return toward values more typical of the early 20th century in recent years. The AO index was computed following Thompson and Wallace (2000) by projecting sea level pressure (SLP) fields onto the leading empirical orthogonal function (EOF) of SLP poleward of 20°N based on all months from January 1958 to April 1997. The monthly SLP data were obtained from the Trenberth SLP data set available from the NCAR Data Distribution Center (<http://dss.ucar.edu/datasets/ds010.1/>) with corrections as described in Trenberth and Paolino (1980).

For the model, data were projected onto the leading EOF of the five (combined) CM2.1 all-forcing scenario runs to form individual model AO indices (green dashed curves in c). Although not shown here, CM2.1 has a highly realistic simulation of the observed AO pattern (Delworth et al. 2005). None of the simulated AO indices have anomalies (from the 1901-1930 reference period) as large as observed for the 1980s and 90s although there is substantial variability in the model AO series. The ensemble mean of the model AO indices (red) shows little evidence for either a trend or for strong positive anomalies late in the 20th century. Gillett et al. (2000) similarly found little evidence for a trend in the AO in their HadCM2 simulations forced by greenhouse gases and sulfate aerosols. However, they also

found that the detection of a global response to these forcings was robust to exclusion of the AO-related warming, a result generally supported by Zwiers and Zhang's (2003) regional-scale analysis.

The impact on surface temperatures of our model's failure to reproduce the observed AO behavior was examined by using the observed AO index to adjust the model's northern Asia temperature series. Linear regressions were first computed between the AO index and the northern Asia time series for each model ensemble member, using annual mean values. The AO influence on the northern Asia temperature series was then removed from each model series using the model's linear regression relationship. The modified 10-yr running-mean ensemble-mean northern Asia temperature index (dashed green curve in Fig. 12 a) is quite similar to the original ensemble-mean index (red curve). This indicates that the AO in the model has minimal impact on multi-decadal variability of the model's northern Asia temperature index. However, the AO does have a significant influence on northern Asia temperatures in the model in general, since if one considers all interannual time scales (i.e., by using unfiltered annual data), the average correlation between the model's AO and northern Asia temperature indices is about 0.43.

An adjusted northern Asia temperature index was obtained by adding to the model's "AO-removed" timeseries the product of the model's ensemble mean regression coefficients and the observed AO index. The resulting curves are labeled "AO-adjusted" in Fig. 12 a, b. The ensemble AO-adjusted index (black dashed in (a)) shows a stronger warming beyond the mid-1980s than does the original unadjusted index (red curve in panel a). Thus, the strong positive observed AO anomalies late in the 20th century, which are mostly absent in the model simulations, appear to account for about half of the ensemble-mean warming discrepancy between the model and observations during this period for the northern Asia temperature index.

Figure 12 (b) shows the individual "AO-adjusted" northern Asia temperature indices from the five CM2.1 all-forcing runs. None of the individual adjusted members have northern Asia warm anomalies as strong as observed in the 1990s according to this comparison, although one of the ensemble members comes fairly close to matching the observed warm anomalies.

The analysis shows that the presence of a positive AO trend in the observations, but not in the model, contributes to the differences in high latitude temperature trends between the model and observations in the late 20th century. In fact, one could argue that if the model had reproduced the observed warming, but without the strong AO anomalies, the "agreement" in surface temperatures for this region would have been for the wrong physical reasons.

8. Summary and Conclusions

Climate models contain our hypotheses about the physical climate system, including how different components of the climate system interact and about the physics required to incorporate important climate forcing agents into climate change experiments. In the experiments described in this paper, two new climate models are forced with our current best estimates of the natural and anthropogenic forcings over the period 1861-2000. We compare model and observed surface temperature variations to check for consistency (or lack of consistency) over various periods. In addition, we use the model as tool with which to probe

and attempt to interpret aspects of the observed surface temperature record. The main findings are as follows:

- 1) The climate model simulations provide additional evidence for the hypothesis that the observed surface temperature warming during the 20th century is too large to have been caused by internal climate variability or natural climate forcings alone. The observed trends exceed model-generated internal variability and are significantly different from the simulated natural-only forcing trends over a wide majority of the global regions tested. There is little evidence that the model simulates too little internal climate variability; in fact, the model appears to simulate excessive internal climate variability on interannual time scales in many regions and at the global scale.
- 2) The historical simulations which include forcing by both natural and anthropogenic agents, or by anthropogenic forcings alone, simulate the observed global mean temperature evolution during 1871-2000 reasonably well. This provides further support for the view that the observed warming is a response of the climate system to the climate forcing agents used in the simulations, with anthropogenic forcings being of particular importance in the late 20th century.
- 3) An exception to the agreement in #2, is that the model shows excessive short-term cooling episodes which coincide with periods of strong volcanic activity. However, the discrepancies for Pinatubo and El Chichon in the late 20th century may be due in large part due to El Niño influences during the post-eruption periods. The causes for the simulated vs observed discrepancies during the volcanically active 1880s requires further investigation.
- 4) Considering the 20th century as a whole, a shortcoming of the model simulations on a regional scale is the tendency for too much warming in the tropics, with too little warming in higher latitudes. The causes of this apparent discrepancy are unclear, but may be partly related to internal variability. For example, in future warming scenarios with large positive radiative forcings, the model warms more in the northern extratropics than in the tropics.
- 5) Considering the last half century (1949-2000), the regional trends in the all-forcings simulations are significantly different from observations over a number of regions. Of particular note is a bias toward too little warming over many high latitude continental regions of the northern hemisphere and parts of the extratropical southern Indian Ocean. These discrepancies appear to be related to the lack of poleward amplification discussed in #4 for the century time scale. There also appears to be too little warming in the tropical Indian Ocean and eastern tropical Pacific, and too much in the tropical North Atlantic during 1949-2000. However for these tropical regions, the time series case studies examining the entire 20th century indicate that the simulated and observed warmings were fairly consistent over the century as a whole, despite discrepancies in the last half century.

- 6) Time series case studies for a number of regions highlight examples of discrepancies and consistencies between the historical simulations and observations. The model agrees with the observations particularly well for the tropical and subtropical Atlantic time series; the results suggest that a century time-scale anthropogenic warming signal is emerging from a background of strong multidecadal variations in these regions. The extratropical north Pacific and high latitude north Atlantic are regions characterized by very strong multi-decadal variations in both models and observations, with long-term trends being difficult to discern. Northern Asia and the Southeastern U.S. emerge as challenging regions to simulate, with the modeled warming being too weak in the former and too strong in the latter. The seasonality of the modeled warming over Northern Asia is also not very realistic. The discrepancies in this region appear due in large part to the lack of a positive trend in the model's AO.

In summary, anthropogenic forcing agents, as opposed to natural forcing agents, are simulated to be the dominant cause of the pronounced warming in the late 20th century. This agrees with a number of previous global modeling studies of natural vs. anthropogenic forcing (e.g., Tett et al. 1999; Stott et al. 2000; Broccoli et al. 2003; Meehl et al. 2004). Establishing the relative contributions of individual forcing agents in CM2 will require further specialized experiments and will be the topic of a future study.

Several factors can contribute to the broad-scale regional discrepancies found in the study, such as too little warming over Northern Asia, Canada, and the southern Indian Ocean, and too much warming over the southeastern U.S. These include 1) errors or omissions in the specified forcing agents, 2) errors in the simulated response to forcing agents; 3) errors in the simulation of internal climate variability; and 4) errors in observed temperature data. These factors, which affect assessments of historical regional and global climate change in general, will require further investigation as the models, specified climate forcing agents, and observational data sets are further refined. In the context of the present model, improved simulations of interannual variability of surface temperature over extratropical land regions and evaluation of the aerosol forcing, including incorporation of indirect aerosol forcing, are important model development goals.

Acknowledgements. We thank Andrew Wittenberg and Kirsten Findell for comments on a draft version of this report.

REFERENCES

- Adams, J. B., M. E. Mann, and C. M. Ammann, 2003: Proxy evidence for an El Nino-like response to volcanic forcing. *Nature*, **426**, 274-278.
- Balaji, V., J. Anderson, I. Held, Z. Liang, M. Winton, and B. Wyman, 2005: FMS: the GFDL Flexible Modeling System: Coupling algorithms for parallel architectures. In preparation.
- Boer, G. J., G. Flato, M. C. Reader, and D. Ramsden, 2000: A transient climate change simulation with greenhouse gas and aerosol forcing: experimental design and comparison with the instrumental record for the twentieth century. *Clim. Dyn.*, **16**, 405-425.
- Broccoli, A. J., K.W. Dixon, T. D. Delworth, T. R. Knutson, R. J. Stouffer, and F. Zeng, 2003: Twentieth-century temperature and precipitation trends in ensemble climate simulations including natural and anthropogenic forcing. *J. Geophys. Res.*, **108**(D24), 4798, doi:10.1029/2003JD003812.
- Eden, C., and T. Jung, 2001: North Atlantic interdecadal variability: oceanic response to the North Atlantic Oscillation (1865-1997). *J. Climate*, **14**, 676-691.
- Delworth, T. L. and co-authors, 2005: GFDL's CM2 global coupled climate models – Part 1: Formulation and simulation characteristics. *J. Climate*, submitted for publication.
- Delworth, T. L., and T. R. Knutson, 2000: Simulation of early 20th Century global warming. *Science*, **287**(5461), 2246-2250.
- Folland, C. K., T. R. Karl, J. R. Christy, R. A. Clarke, G. V. Gruza, J. Jouzel, M. E. Mann, J. Oerlemans, M. J. Salinger, and S.-W. Wang, 2001: Observed climate variability and change. In: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson (eds.)]. Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, 881 pp.
- GFDL Global Atmospheric Model Development Team, 2004: The new GFDL global atmosphere and land model AM2/LM2: Evaluation with prescribed SST simulations. *J. Climate*, **17**, 4641-4673.
- Gillett, N. P., M. F. Wehner, S. F. B. Tett, and A. J. Weaver, 2004: Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. *Geophys. Res. Lett.*, **31**, L14201, doi: 10.1029/2004GL020111.
- Gnanadesikan, A., and Co-authors, 2005: GFDL's CM2 Global Coupled Climate Models: Part 3. The baseline ocean simulation. *J. Climate*, submitted for publication.

- Goldenberg, S. B., C. W. Landsea, A.M. Mesta-Nuñez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: causes and implications. *Science*, **293**, 474-479.
- Griffies, S. M., et al., 2005: Formulation of an ocean model for global climate simulations. GFDL Ocean Group Technical Report No. 6.
- Haywood, J. M., R. J. Stouffer, R. T. Wetherald, S. Manabe, and V. Ramaswamy, 1997: Transient response of a coupled model to estimated changes in greenhouse gas and sulfate concentrations. *Geophys. Res. Lett.*, **24**(11), 1335-1338.
- Hansen, J. A., M. Sato, and R. Ruedy, 1997: Radiative forcing and climate response. *J. Geophys. Res.*, **102**, 6831-6864.
- Hansen, J. E., and coauthors, 2002: Climate forcing in Goddard Institute for Space Studies SI2000 simulations, *J. Geophys. Res.*, **107**, NO. D18, doi:10.1029/2001JD001143.
- Hansen, J. E., L. Nazarenko, R. Ruedy, M. Sato, J. Willis, A. Del Genio, D. Koch, A. Lacis, K. Lo, S. Menon, T. Novakov, J. Perlwitz, G. Russell, G. A. Schmidt, and N. Tausnev, 2005: Earth's energy imbalance: confirmation and implications. *Scienceexpress*, doi: 10.1126/science.1110252 (www.scienceexpress.org).
- Hegerl, G. C., K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Clim. Dyn.*, **13**, 613-634, 1997.
- Hoerling, M. P., J. W. Hurrell, T. Xu, G. T. Bates, and A. S. Phillips, 2004: Twentieth century North Atlantic climate change. Part II: understanding the effect of Indian Ocean warming. *Clim. Dyn.*, **23**, 391-405.
- Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676-679.
- Hurt, G. C., S. Frolking, M. G. Fearon, B. Moore III, E. Shevliakova, S. Malyshev, S. W. Pacala, and R. A. Houghton, 2005: The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood harvest activity, and resulting secondary lands. *Global Change Biology*. Submitted for publication.
- IPCC, 2001. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp.

- Jones, P.D., K. R. Briffa, and F. H. Schweingruber, 1995: Tree-ring evidence of the widespread effects of explosive volcanic eruptions, *Geophys. Res. Lett.*, **22**, 1333-1336.
- Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *J. Climate*, **16**, 206-223.
- Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, E. B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner, 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.*, **106**(D4), 3371-3380.
- Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991, *J. Geophys. Res.*, **103**, 18,567-18,589.
- Karoly, D. J., K. Braganza, P. A. Stott, J. M. Arblaster, G. A. Meehl, A. J. Broccoli, and K. W. Dixon, 2003: Detection of a human influence on North American climate, *Science*, 302, 1200-1203.
- Knutson, T. R., T. L. Delworth, K. W. Dixon, and R. J. Stouffer, 1999: Model assessment of regional surface temperature trends (1949-1997). *J. Geophys. Res.*, **104**(D24), 30,981-30,996.
- Lean, J., J. Beer, and R. S. Bradley, 1995: Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophys. Res. Lett.*, **22**, 3195-3198.
- Lin, S.-J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**(10), 2293-2307.
- Meehl, G. M., W. M. Washington, C. M. Amman, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi, 2004: Combinations of natural and anthropogenic forcings in twentieth-century climate. *J. Climate*, **17**, 3721-3727.
- Milly, P. C. D., and A. B. Shmakin, 2002: Global modeling of land water and energy balances. Part I: The land dynamics (LaD) model. *Journal of Hydrometeorology*, **3**(3), 283-299.
- Mitchell, J. F. B., D. J. Karoly, G. C. Hegerl, F. W. Zwiers, M. R. Allen, and J. Marengo, 2001: Detection of climate change and attribution of causes. In: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson (eds.)]. Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, 881 pp.

- Parker, D. E., C. K. Folland, and M. Jackson, 1995: Marine surface temperature: Observed variations and data requirements. *Clim. Change*, **31**, 559-600.
- Ramaswamy, V., O. Boucher, J. Haigh, D. Hauglustaine, J. Haywood, G. Myhre, T. Nakajima, G. Y. Shi, S. Solomon, 2001: Radiative forcing of climate change. In: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson (eds.)]. Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, 881 pp.
- Ramaswamy, V., and C-T. Chen, 1997: Linear additivity of climate response for combined albedo and greenhouse perturbations. *Geophys. Res. Lett.*, **24**(5), 567-570.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, **108** (D14), 4407, doi:10.1029/2002JD002670.
- Santer, B. D., T. M. L. Wigley, D. Doutriaux, J. S. Boyle, J. E. Hansen, P. D. Jones, G. A. Meehl, E. Roeckner, S. Sengupta, and K. E. Taylor, 2001: Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. *J. Geophys. Res.*, **106** (D22), 28,033-23,059.
- Sato, M., J. Hansen, M. P. McCormick, and J. Pollack, 1993: Stratospheric Aerosol Optical Depth, 1850-1990, *J. Geophys. Res.*, **98**, 22,987-22,994.
- Smith, T. M., and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854-1997), *J. Clim.*, **16**, 1495-1510.
- Soden, B. J., R. T. Wetherald, G. L. Stenchikov, and A. Robock, 2002: Global cooling after the eruption of Mount Pinatubo: A test of climate feedback by water vapor. *Science*, **296**(5568), 727-730.
- Stenchikov, G., I. Kirchner, A. Robock, H. Graf, J. Antuna, R. G. Grainger, A. Lambert, and L. Thomason, 1998: Radiative forcing from the 1991 Mount Pinatubo volcanic eruption, *J. Geophys. Res.*, **103**, 13,837-13,857.
- Stott, P. A., 2003: Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.*, **30**, 1728, doi:10.1029/2003GL017324.
- Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133-2137.

- Stouffer, R. J., G. Hegerl, and S. Tett, 2000: A comparison of surface air temperature variability in three 1000-yr coupled ocean-atmosphere model integrations. *J. Climate*, **13**(3), 513-537.
- Stouffer, R. J., A. J. Weaver, and M. Eby, 2004: A method for obtaining pre-twentieth century initial conditions for use in climate change studies. *Clim. Dyn.*, **23**, 327-339.
- Stouffer, R. J., and Co-authors, 2005: GFDL's CM2 Global Coupled Climate Models – Part 4: Idealized climate change. *J. Climate*, submitted for publication.
- Tett, S. F. B., P. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth-century temperature change near the Earth's surface. *Nature*, **399**, 569-572.
- Thompson, D.W. J. and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000-1016.
- Thompson, D. W. J., J. M. Wallace, and G. C. Hegerl, 2000: Annular modes in the extratropical circulation. Part II: Trends. *J. Climate*, **13**, 1018-1036.
- Trenberth, K. E., and D. A. Paolino, Jr., 1980: The Northern Hemisphere sea-level pressure data set: Trends, errors, and discontinuities. *Mon. Wea. Rev.*, **108**, 855-872.
- Winton, M., 2000: A reformulated three-layer sea ice model. *Journal of Atmospheric & Oceanic Technology*, **17**(4), 525-531.
- Wittenberg, A. T., A. Rosati, N.-C. Lau, and J. J. Ploshay, 2005: GFDL's CM2 global coupled climate models, Part 4: Tropical Pacific climate and ENSO. Submitted to *J. Climate*.
- Zwiers, F. W. and X. Zhang, 2003: Toward regional-scale climate change detection. *J. Climate*, **16**, 793-797.

Figure Captions

Fig. 1. Time series of global mean reference temperature (T_{ref}) from a) CM2.0 control run and b) CM2.1 control run (black curves). The red curves in each diagram are the observed surface temperature anomalies from HadCRUT2v (see text). The observations are offset by an arbitrary constant and are presented only as a reference trend comparison for the model results. The straight line segments connecting “x” marks depict different segments of the control run, including the segments corresponding to the historical forcing runs. The line segments superimposed on the time series illustrate the periods used for detrending of the post-“spin-up” segments of the control run. The “spin-up” segments correspond to preliminary parts of the integration that are not analyzed. Note that the two diagrams have time axes of different length.

Fig. 2. Standard deviations of annual mean surface temperature (SST over ocean, surface air temperature over land) for a) observations from HadCRUT2v (1949-2003); b) CM2.0 control run; c) CM2.1 control run; and d) modified CM2.1 control run (see text). Model segments are 100 years in length. Contour interval: 0.1°C . All time series were detrended prior to computing the standard deviations.

Fig. 3. Global annual mean temperature time series for the observations (HadCRUT2v, black) and model historical forcing runs: a) CM2.0 all forcings; b) CM2.1 all forcings; c) CM2.1 all forcings vs natural-only forcings plus anthropogenic-only forcings; d) CM2.1 natural forcings only; and e) CM2.1 anthropogenic forcings only. In a, b, d, and e, thick red curves are model ensemble means, and green dashed curves are individual ensemble members; in c the all-forcings ensemble mean is red and the sum of the natural-only ensemble and the anthropogenic-only ensemble is dashed blue. All curves are referenced to the period 1881-1920. Model results use SST and T_{ref} over ocean and land regions, and are masked-out during periods with no observations.

Fig. 4. Variance spectra of global mean surface temperature series for observations (1871-2000; red curves) and for 130-yr segments from various all-forcing historical experiments for: a) CM2.0 and b) CM2.1. The models’ ensemble-mean spectra are depicted by medium solid black lines, and are computed as the mean of the spectra of the individual ensemble members (thin lines). The raw spectra were smoothed using a non-overlapping boxcar window of width 3 calculable frequencies. All series were detrended prior to spectral calculations. Model time series have been constructed using the observed data mask.

Fig. 5. (a) Observed surface temperature trends (1949-2000) in $^{\circ}\text{C}$ per 100yr. (b) Black shading indicates regions where the observed trend is within the 5th-95th percentile range of the simulated trends from 23 non-overlapping 52-yr segments of the control run. Color shaded regions thus show where observed trends are unusual in comparison to model internal variability according to this measure. (c,e,g) Simulated surface temperature trends for 1949-2000 based on the ensemble mean ($n=8$) of the CM2.0 and CM2.1 all-forcing (c), natural-only forcing (e), or anthropogenic-only forcing (g) historical runs. (d,f,h) Black regions indicate where the null hypothesis that the observed and ensemble-mean simulated

trends are equal cannot be rejected at the 0.1 level using a 2-sided, 2-sample modified t-test (see text). Thus, color-shaded regions in (d,f,h) show model minus observed differences in those regions where these differences are significant at the 0.1 level for the all-forcing (d), natural-only forcing (f), or anthropogenic-only forcing (h) historical runs.

Fig. 6. As in Fig. 5 but for separate 3-month seasons: a) December through February (DJF); b) March through May (MAM); c) June through August (JJA); and d) September through November (SON). For each season, the same 4-panel display format as for Fig. 5 (a-d) is repeated (i-iv).

Fig. 7. Maps of 100-yr trends of surface temperature (1901-2000) in $^{\circ}\text{C}$ per 100yr for (a, b) observations; (c, d, e) CM2.0 all-forcing historical runs ($n=3$ ensemble mean); and (f, g, h) CM2.1 all-forcing historical runs ($n=5$ ensemble mean). The left column (labelled “No Masking”) uses all model data, with no missing data masking. The center and right columns (labeled “5% availability criterion” and “30% availability criterion”) have trends plotted only for grid points where the observations have at least 5% (or 30%) data availability of monthly mean anomalies for each of five 20 years periods (1901-1920, 1921-1940, ..., 1981-2000). The model fields are masked out during observed missing data periods for the trend calculations shown in these columns.

Fig. 8. Map showing the areas used to create various area-averaged time series for the case studies in Figs. 9-11. The percent of global area covered for each regions is: Southeast U.S: 0.4%; Indian Ocean/Western Pacific Warm Pool: 8.6%; Subtropical North Atlantic: 3.3%; Northern Asia: 3.7%; North Atlantic: 0.8%; North Pacific: 2.6%; Tropical East Pacific: 7.4%; Tropical North Atlantic: 1.4%; Tropics: 34.2%; Northern Extratropics: 32.9%; Southern Extratropics: 32.9%.

Fig. 9. Ten-yr running-mean area-averaged time series of surface temperature anomalies ($^{\circ}\text{C}$) for observations and models for various regions: (a-c) Global mean; (d-f) northern hemisphere extratropics (20°N - 90°N); (g-i) southern hemisphere extratropics (20°S - 90°S); and (j-l) tropics (20°N - 20°S). The left column (a, d, g, j) shows results for the CM2.0 all-forcing historical runs ($n=3$) through 2000 and observations through 2004. The middle column shows results for CM2.1 all-forcing runs ($n=5$) and observations. The right column shows observations through 2000, with ± 2 -standard deviation ranges (shading) based on uncertainties due to missing data as estimated by sampling the model runs. The red, blue, and green curves in the right-column diagrams are ensemble mean results for the CM2.1 all-forcing ($n=5$), natural-only ($n=3$), and anthropogenic-only ($n=3$) forcing historical runs. Model data was masked according to observed data coverage, and 1881-1920 was used as the reference period.

Fig. 10. As in Fig. 9, but for the following regions: (a-c) Indian Ocean/Western Pacific Warm Pool; (d-f) tropical eastern Pacific; (g-i) tropical North Atlantic; and (j-l) subtropical North Atlantic. See Fig. 8 for illustrated location of specific regions.

Fig. 11. As in Fig. 9, but for the following regions: (a-c) North Atlantic (45-60°N); (d-f) North Pacific (25-45°N); (g-i) northern Asia; and (j-l) southeastern U.S. See Fig. 8 for illustrated location of specific regions.

Fig. 12. Time series of 10-yr running mean northern Asia temperature index (Fig. 8) for observations (solid black) and the CM2.1 historical runs. a) Model ensemble mean (red), model ensemble mean with model AO contribution removed (green dashed) and model ensemble mean with model AO contribution replaced with an adjustment according to observed AO variability (black dashed). See text for details. b) Observed northern Asia index (thick black) and “AO-adjusted” indices for individual CM2.1 ensemble members (thin colored).

TABLE CAPTION:

Table 1. Standard deviations of annual-mean or 10-yr running mean global-mean surface temperature time series (1871-2000) for observations and “All-Forcing” model historical runs (unit: degrees Celsius). For the “Detrended” columns, a least-squares linear trend has been removed from the data prior to computing the standard deviations. The ratio is obtained by dividing the average model standard deviation by the observed standard deviation.

Table 1. Standard deviations of annual-mean or 10-yr running mean global-mean surface temperature time series (1871-2000) for observations and “All-Forcing” model historical runs (unit: degrees Celsius). For the “Detrended” columns, a least-squares linear trend has been removed from the data prior to computing the standard deviations. The ratio is obtained by dividing the average model standard deviation by the observed standard deviation.

		Unfiltered annual means	Detrended annual means	Detrended 10-yr running means
	Observed	0.222	0.128	0.082
	CM2.0 h1	0.218	0.167	0.095
	CM2.0 h2	0.215	0.165	0.090
	CM2.0 h3	0.212	0.148	0.078
	CM2.0 Average	0.215	0.160	0.088
	Ratio: CM2.0 to Obs	0.968	1.250	1.069
	CM2.1 H1	0.286	0.205	0.105
	CM2.1 H2	0.248	0.170	0.087
	CM2.1 H3	0.284	0.184	0.111
	CM2.1 H4	0.204	0.159	0.071
	CM2.1 H5	0.253	0.167	0.081
	CM2.1 Average	0.255	0.177	0.091
	Ratio: CM2.1 to Obs	1.149	1.383	1.110

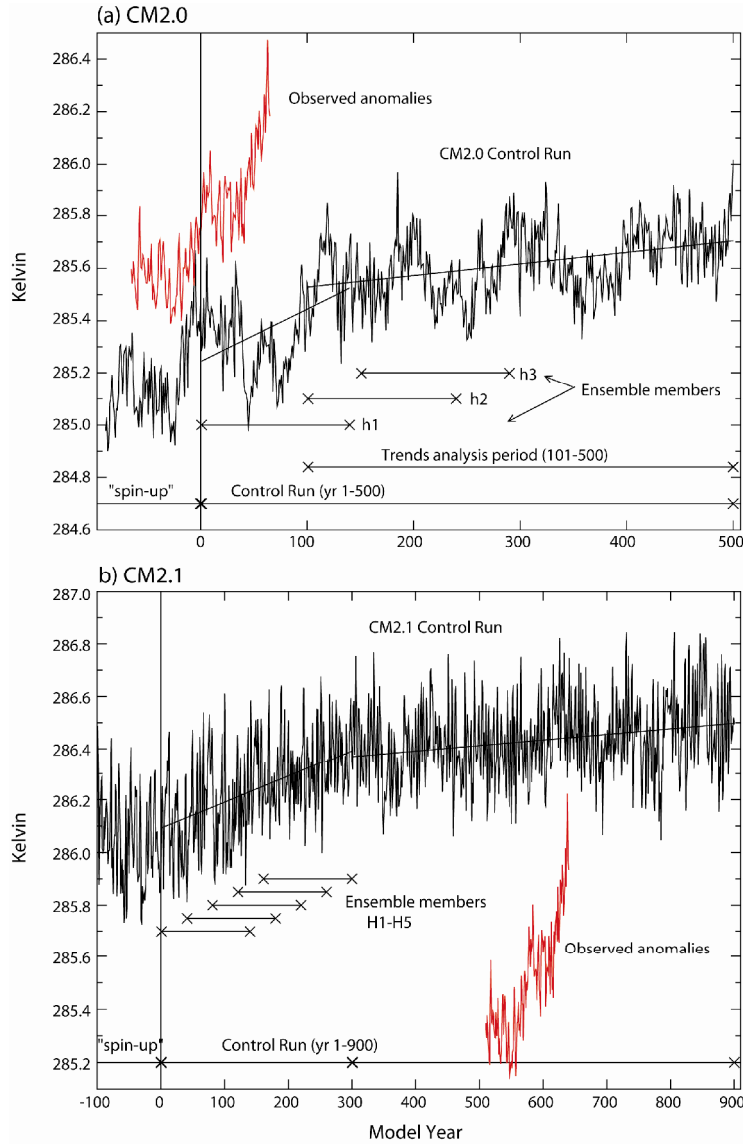


Fig. 1. Time series of global mean reference temperature (T_{ref}) from a) CM2.0 control run and b) CM2.1 control run (black curves). The red curves in each diagram are the observed surface temperature anomalies from HadCRUT2v (see text). The observations are offset by an arbitrary constant and are presented only as a reference trend comparison for the model results. The straight line segments connecting “x” marks depict different segments of the control run, including the segments corresponding to the historical forcing runs. The line segments superimposed on the time series illustrate the periods used for detrending of the post-“spin-up” segments of the control run. The “spin-up” segments correspond to preliminary parts of the integration that are not analyzed. Note that the two diagrams have time axes of different length.

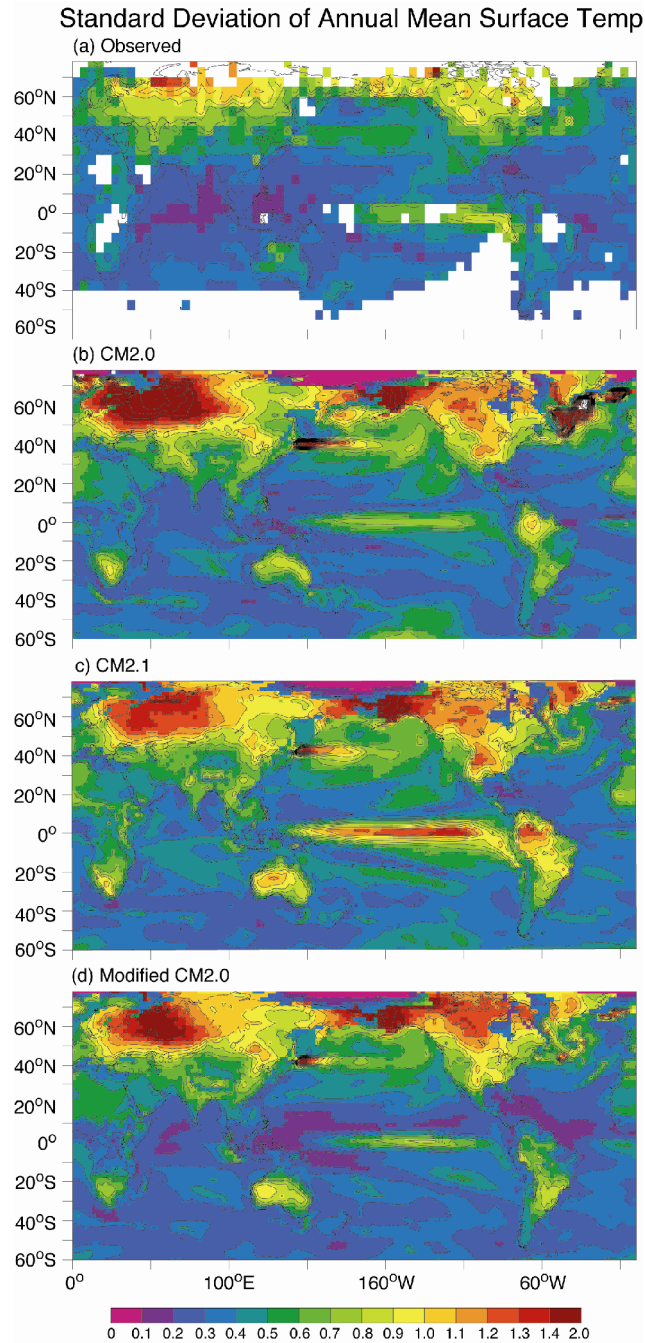


Fig. 2. Standard deviations of annual mean surface temperature (SST over ocean, surface air temperature over land) for a) observations from HadCRUT2v (1949-2003); b) CM2.0 control run; c) CM2.1 control run; and d) modified CM2.1 control run (see text). Model segments are 100 years in length. Contour interval: 0.1°C . All time series were detrended prior to computing the standard deviations.

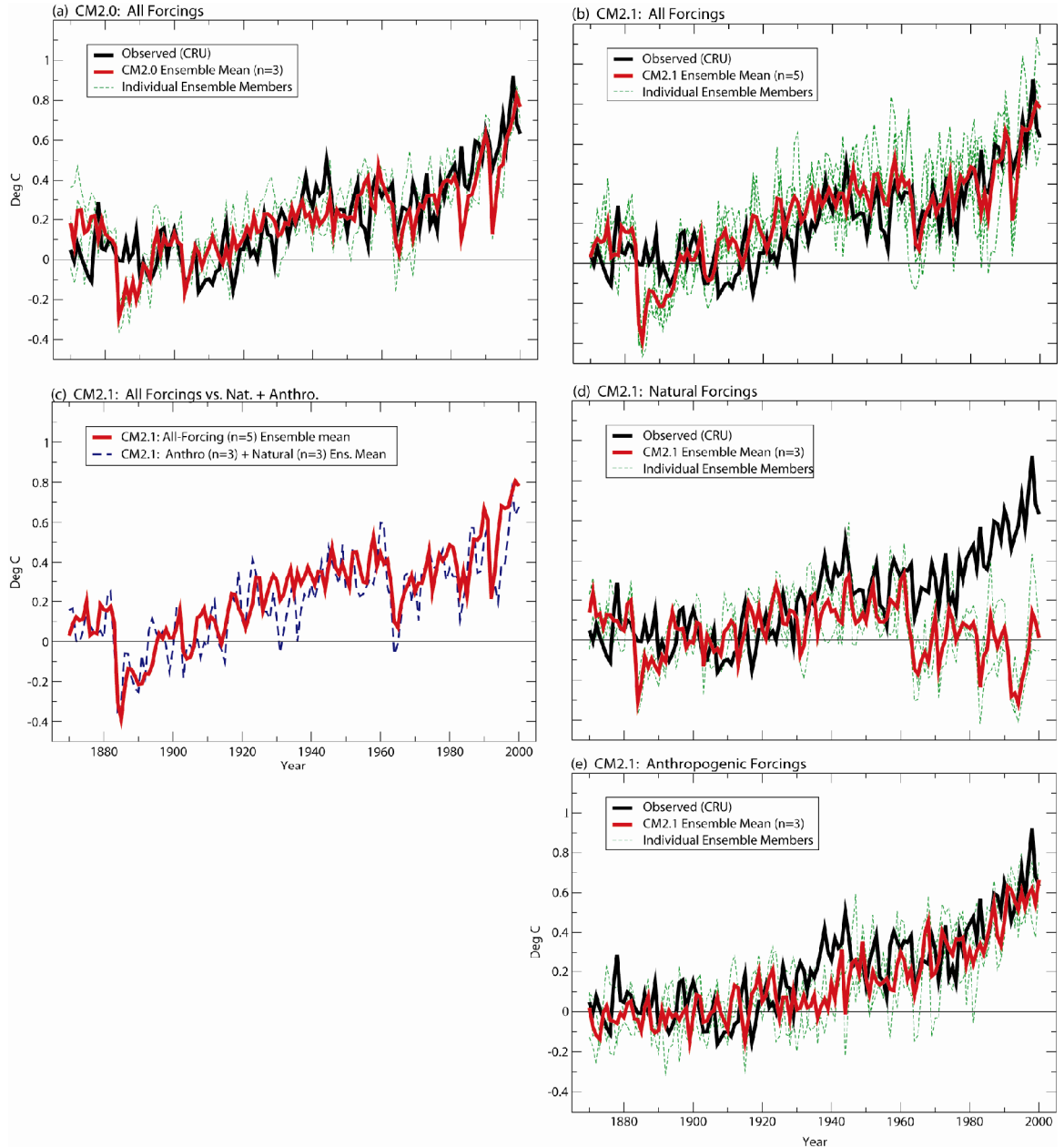


Fig. 3. Global annual mean temperature time series for the observations (HadCRUT2v, black) and model historical forcing runs: a) CM2.0 all forcings; b) CM2.1 all forcings; c) CM2.1 all forcings vs natural-only forcings plus anthropogenic-only forcing; d) CM2.1 natural forcings only; and e) CM2.1 anthropogenic forcings only. In a, b, d, and e, thick red curves are model ensemble means, and green dashed curves are individual ensemble members; in c the all-forcings ensemble mean is red and the sum of the natural-only ensemble and the anthropogenic-only ensemble is dashed blue. All curves are referenced to the period 1881-1920. Model results use SST and T_{ref} over ocean and land regions, and are masked-out during periods with no observations.

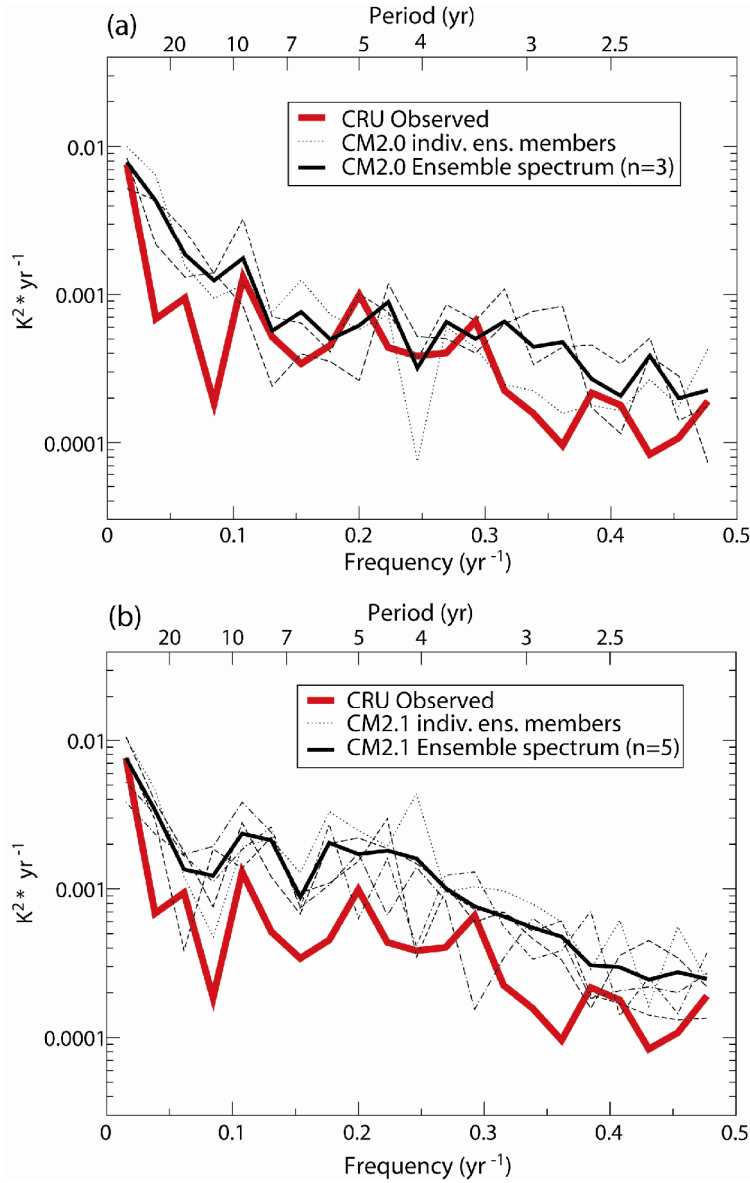


Fig. 4. Variance spectra of global mean surface temperature series for observations (1871-2000; red curves) and for 130-yr segments from various all-forcing historical experiments for: a) CM2.0 and b) CM2.1. The models' ensemble-mean spectra are depicted by medium solid black lines, and are computed as the mean of the spectra of the individual ensemble members (thin lines). The raw spectra were smoothed using a non-overlapping boxcar window of width 3 calculable frequencies. All series were detrended prior to spectral calculations. Model time series have been constructed using the observed data mask.

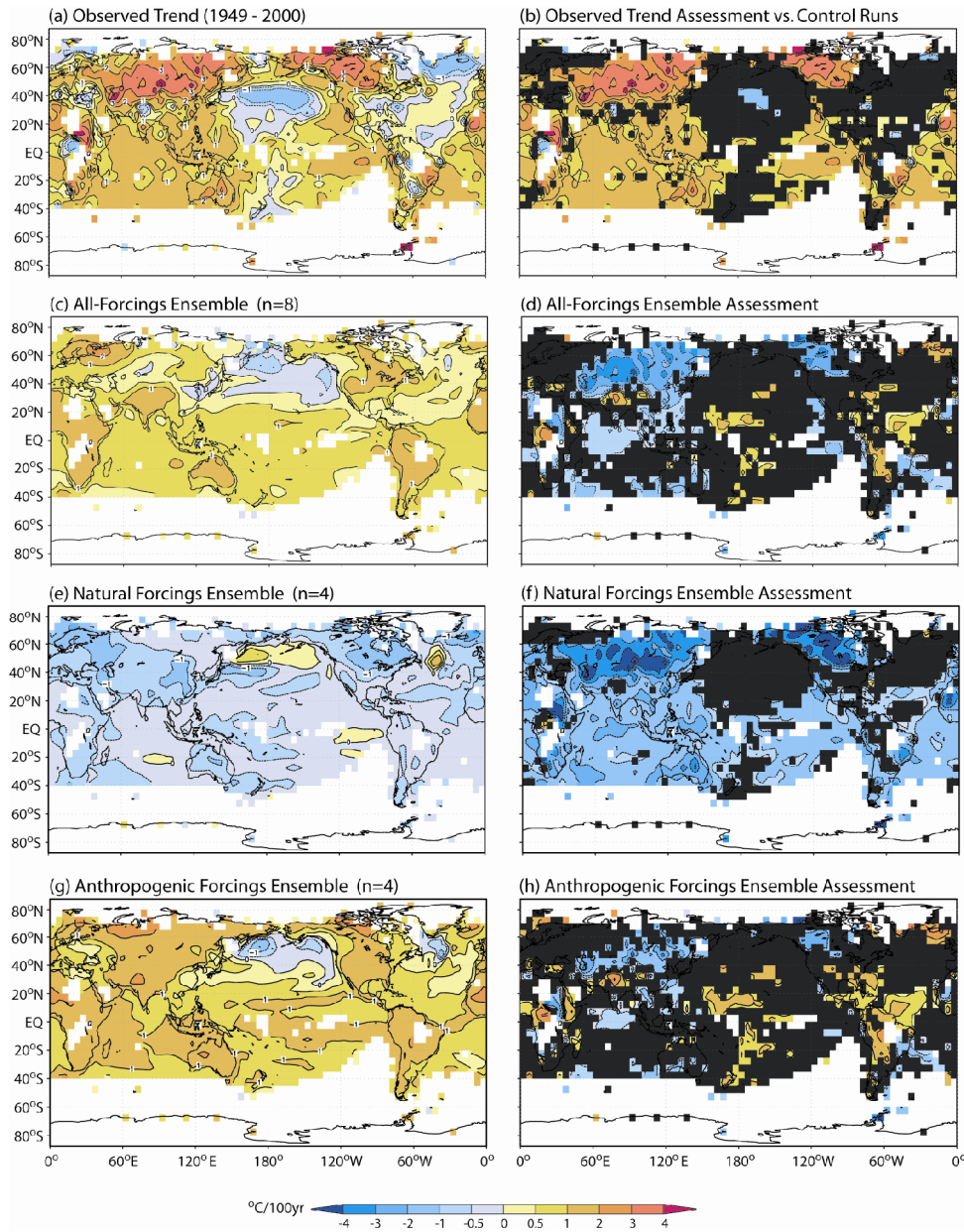


Fig. 5. (a) Observed surface temperature trends (1949-2000) in °C per 100yr. (b) Black shading indicates regions where the observed trend is within the 5th-95th percentile range of the simulated trends from 23 non-overlapping 52-yr segments of the control run. Color shaded regions thus show where observed trends are unusual in comparison to model internal variability according to this measure. (c,e,g) Simulated surface temperature trends for 1949-2000 based on the ensemble mean (n=8) of the CM2.0 and CM2.1 all-forcing (c), natural-only forcing (e), or anthropogenic-only forcing (g) historical runs. (d,f,h) Black regions indicate where the null hypothesis that the observed and ensemble-mean simulated trends are equal cannot be rejected at the 0.1 level using a 2-sided, 2-sample modified t-test (see text). Thus, color-shaded regions in (d,f,h) show model minus observed differences in those regions where these differences are significant at the 0.1 level for the all-forcing (d), natural-only forcing (f), or anthropogenic-only forcing (h) historical runs.

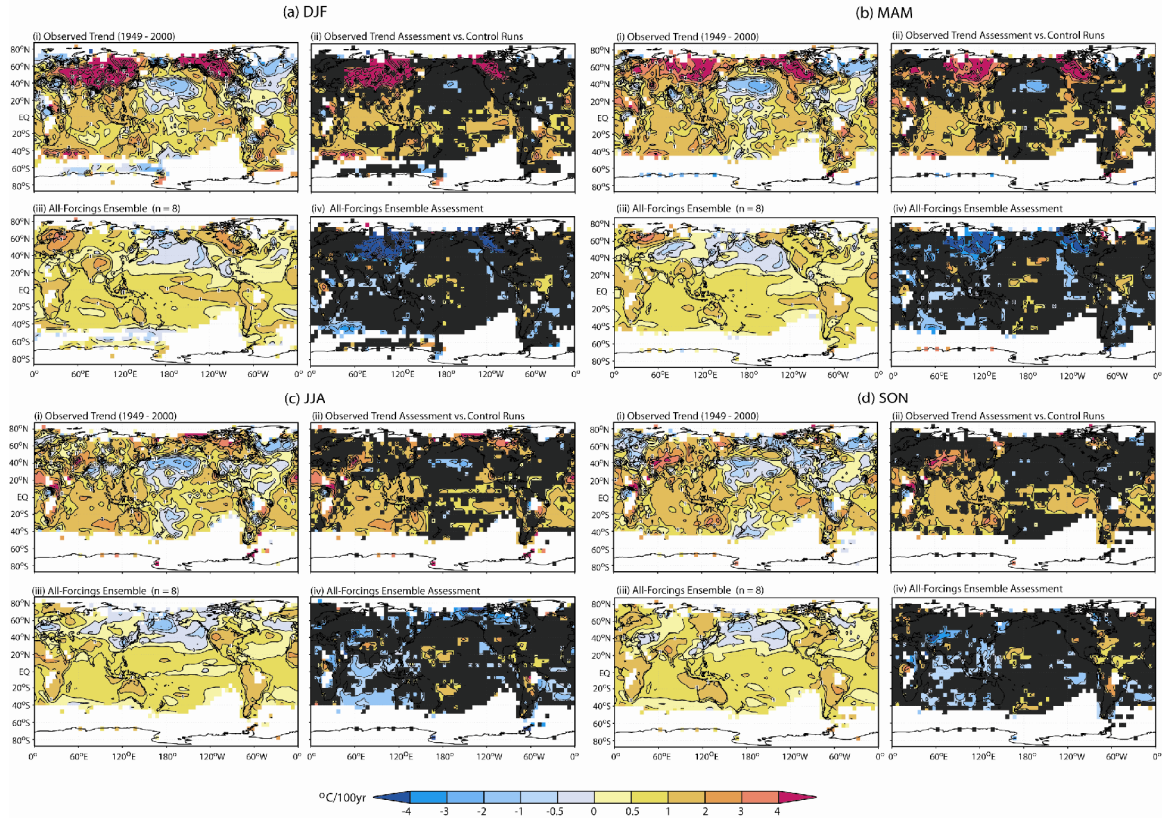


Fig. 6. As in Fig. 5 but for separate 3-month seasons: a) December through February (DJF); b) March through May (MAM); c) June through August (JJA); and d) September through November (SON). For each season, the same 4-panel display format as for Fig. 5 (a-d) is repeated (i-iv).

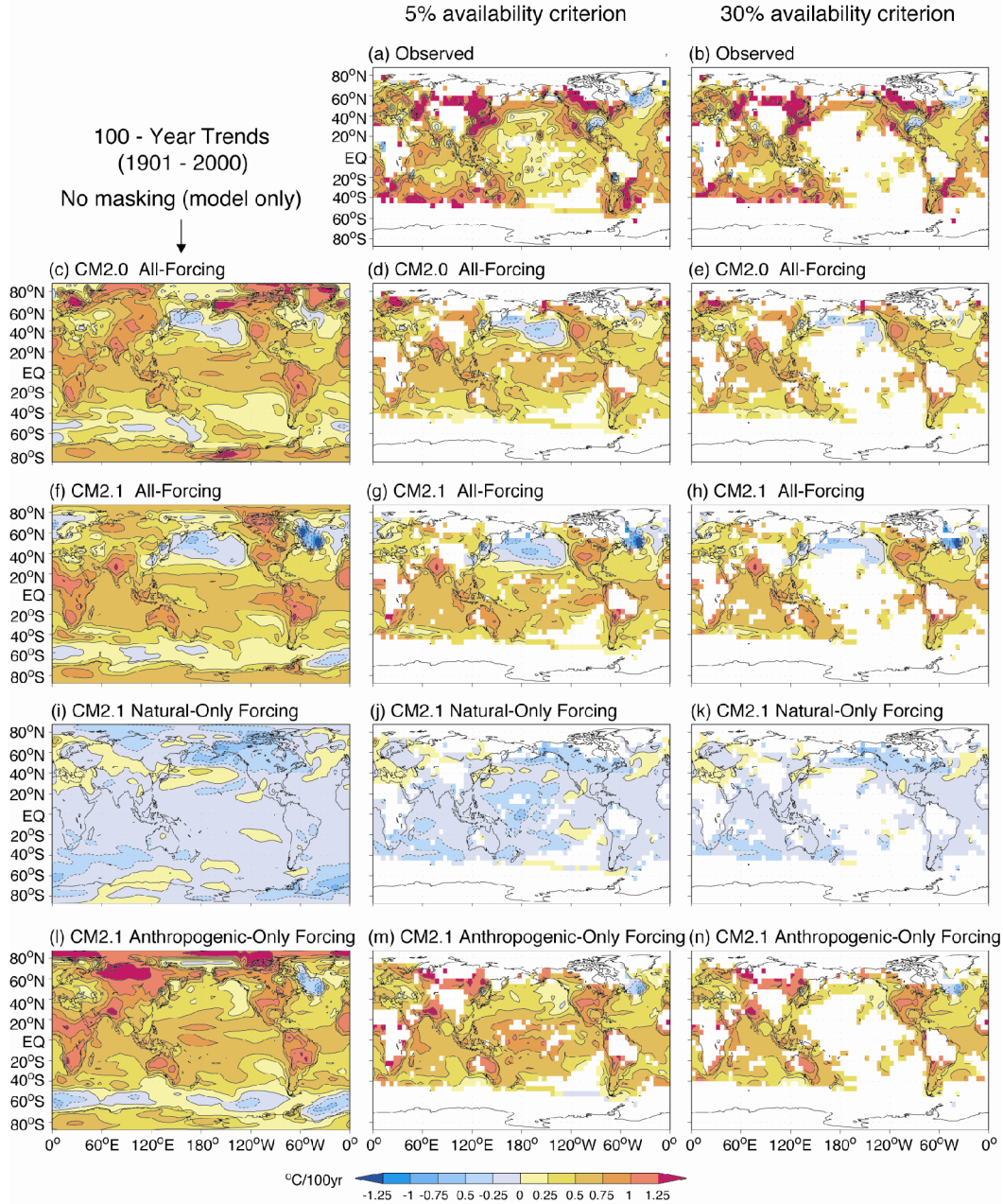


Fig. 7. Maps of 100-yr trends of surface temperature (1901-2000) in $^{\circ}\text{C}$ per 100yr for (a, b) observations; (c, d, e) CM2.0 all-forcing historical runs ($n=3$ ensemble mean); and (f, g, h) CM2.1 all-forcing historical runs ($n=5$ ensemble mean). The left column (labelled “No Masking”) uses all model data, with no missing data masking. The center and right columns (labeled “5% availability criterion” and “30% availability criterion”) have trends plotted only for grid points where the observations have at least 5% (or 30%) data availability of monthly mean anomalies for each of five 20 years periods (1901-1920, 1921-1940, ..., 1981-2000). The model fields are masked out during observed missing data periods for the trend calculations shown in these columns.

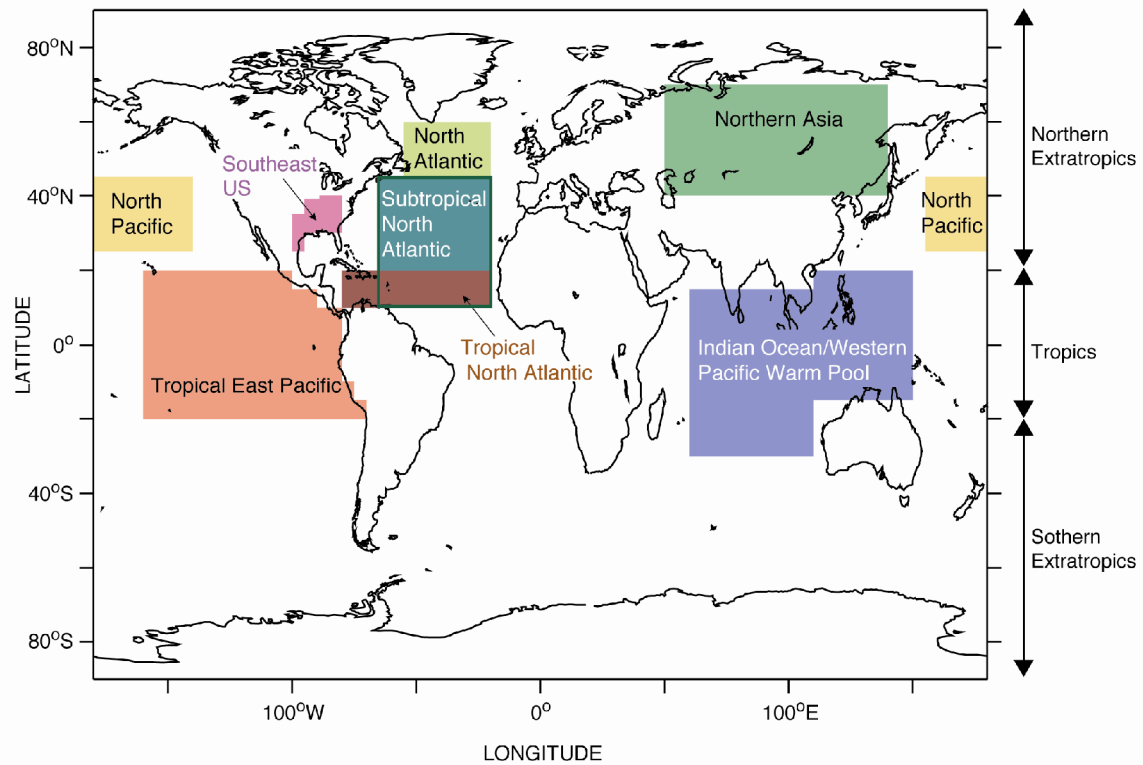


Fig. 8. Map showing the areas used to create various area-averaged time series for the case studies in Figs. 9-11. The percent of global area covered for each regions is: Southeast U.S: 0.4%; Indian Ocean/Western Pacific Warm Pool: 8.6%; Subtropical North Atlantic: 3.3%; Northern Asia: 3.7%; North Atlantic: 0.8%; North Pacific: 2.6%; Tropical East Pacific: 7.4%; Tropical North Atlantic: 1.4%; Tropics: 34.2%; Northern Extratropics: 32.9%; Southern Extratropics: 32.9%.

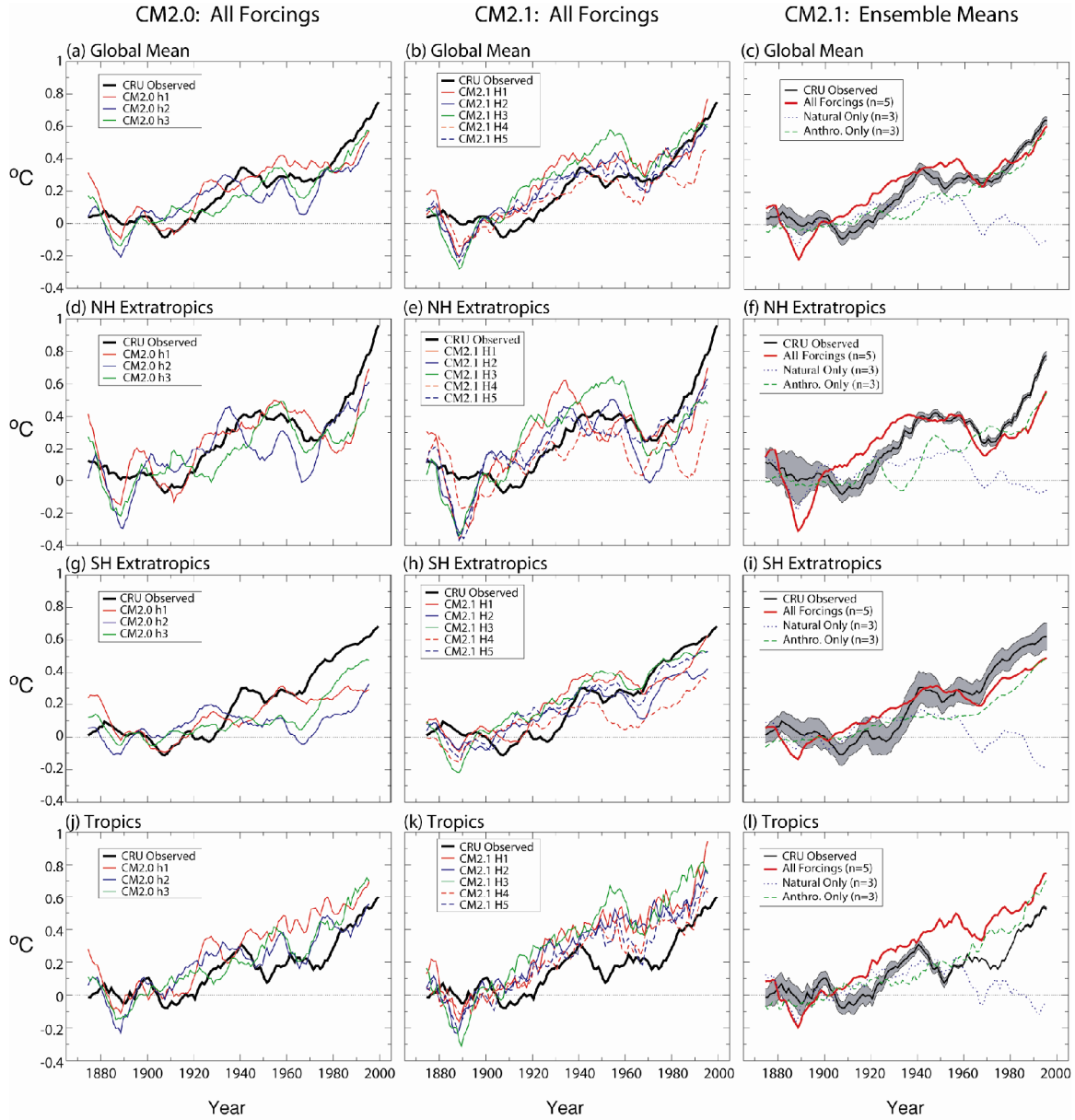


Fig. 9. Ten-yr running-mean area-averaged time series of surface temperature anomalies ($^{\circ}\text{C}$) for observations and models for various regions: (a-c) Global mean; (d-f) northern hemisphere extratropics ($20\text{--}90^{\circ}\text{N}$); (g-i) southern hemisphere extratropics ($20\text{--}90^{\circ}\text{S}$); and (j-l) tropics ($20^{\circ}\text{N}\text{--}20^{\circ}\text{S}$). The left column (a, d, g, j) shows results for the CM2.0 all-forcing historical runs ($n=3$) through 2000 and observations through 2004. The middle column shows results for CM2.1 all-forcing runs ($n=5$) and observations. The right column shows observations through 2000, with ± 2 -standard deviation ranges (shading) based on uncertainties due to missing data as estimated by sampling the model runs. The red, blue, and green curves in the right-column diagrams are ensemble mean results for the CM2.1 all-forcing ($n=5$), natural-only ($n=3$), and anthropogenic-only ($n=3$) forcing historical runs. Model data was masked according to observed data coverage, and 1881-1920 was used as the reference period.

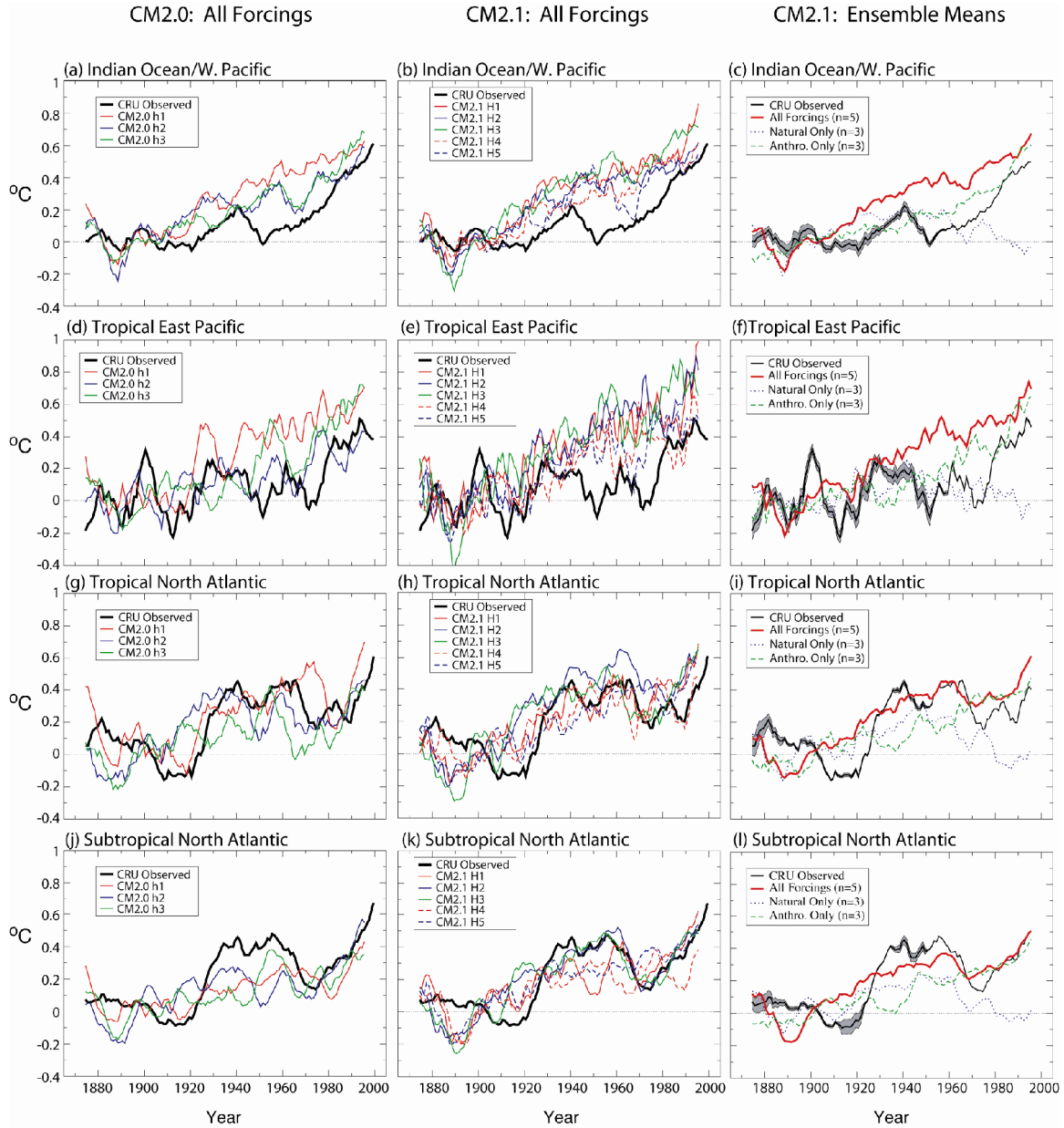


Fig. 10. As in Fig. 9, but for the following regions: (a-c) Indian Ocean/Western Pacific Warm Pool; (d-f) tropical eastern Pacific; (g-i) tropical North Atlantic; and (j-l) subtropical North Atlantic. See Fig. 8 for illustrated location of specific regions.

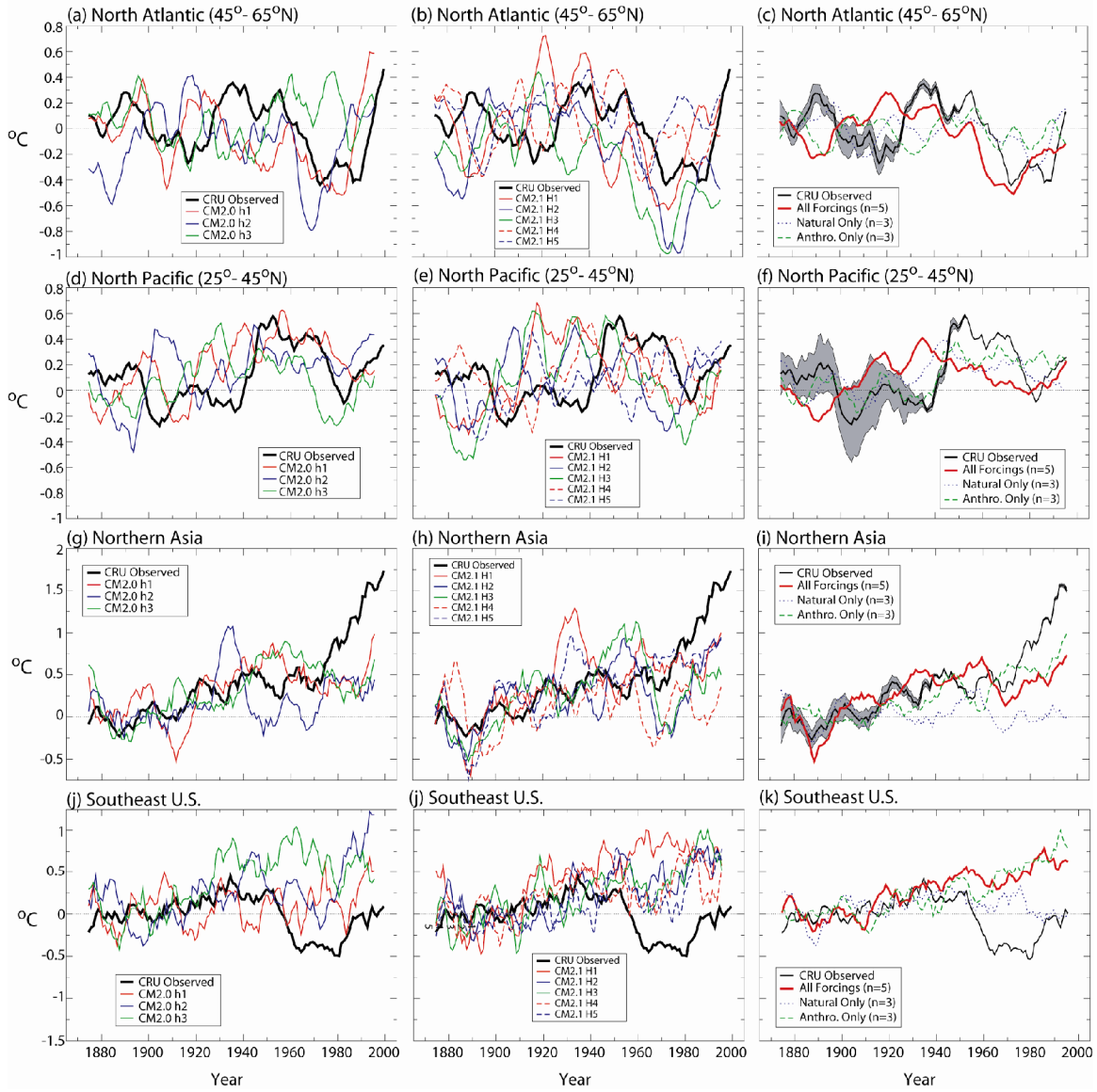


Fig. 11. As in Fig. 9, but for the following regions: (a-c) North Atlantic (45-60°N); (d-f) North Pacific (25-45°N); (g-i) northern Asia; and (j-l) southeastern U.S. See Fig. 8 for illustrated location of specific regions.

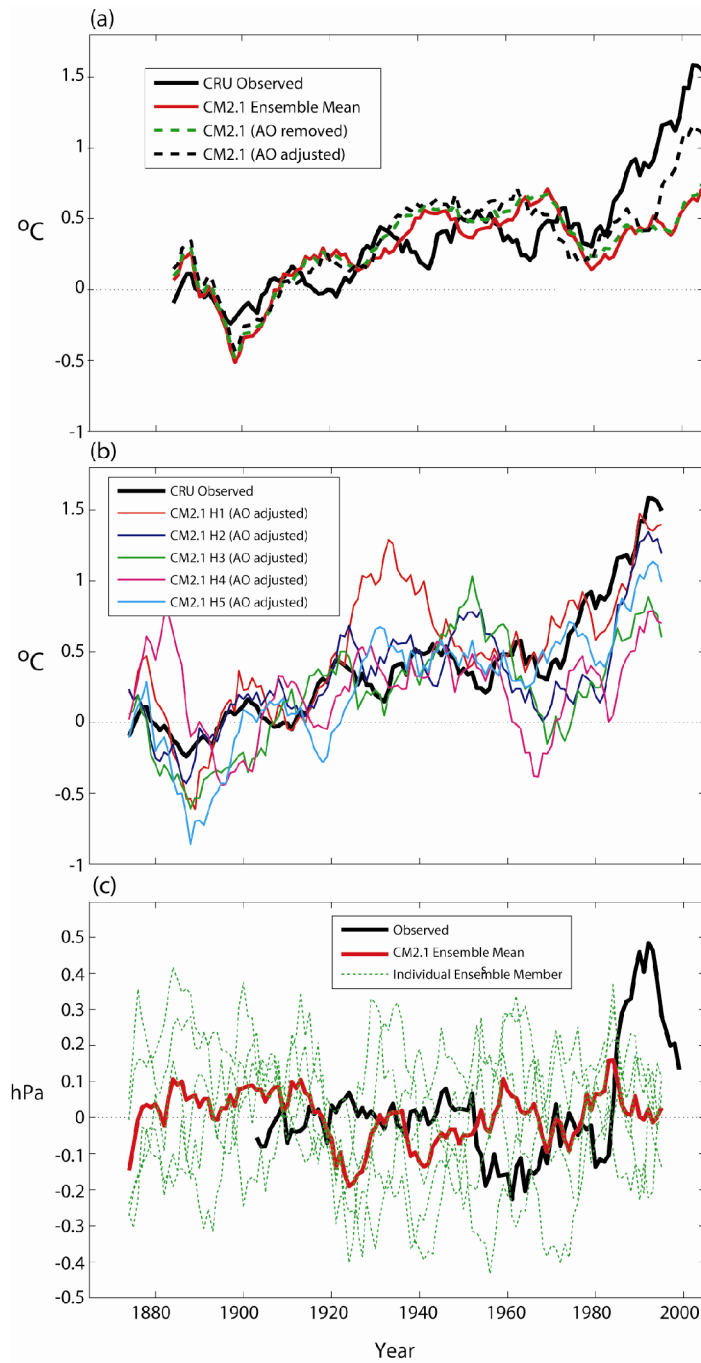


Fig. 12. Time series of 10-yr running mean northern Asia temperature index (Fig. 8) for observations (solid black) and the CM2.1 historical runs. a) Model ensemble mean (red), model ensemble mean with model AO contribution removed (green dashed) and model ensemble mean with model AO contribution replaced with an adjustment according to observed AO variability (black dashed). See text for details. b) Observed northern Asia index (thick black) and “AO-adjusted” indices for individual CM2.1 ensemble members (thin colored).